Framework

AWS Well-Architected Framework

AWS Well-Architected Framework: Framework

Table of Contents

AWS Well-Architected Framework

Data di pubblicazione: 27 giugno 2024 (Revisioni del documento)

Il AWS Well-Architected Framework ti aiuta a comprendere i pro e i contro delle decisioni che prendi durante la creazione di sistemi. AWS Utilizzando il Framework, scoprirai le best practice architetturali per progettare e gestire sistemi affidabili, sicuri, efficienti, convenienti e sostenibili nel cloud.

Introduzione

Il AWS Well-Architected Framework ti aiuta a comprendere i pro e i contro delle decisioni che prendi durante la creazione di sistemi. AWS Utilizzando il Framework, scoprirai le best practice architetturali per progettare e gestire carichi di lavoro affidabili, sicuri, efficienti, convenienti e sostenibili nell' Cloud AWS. Offre un metodo per misurare con coerenza le proprie architetture rispetto alle best practice e individuare le aree di miglioramento. Il processo di revisione di un'architettura è una conversazione costruttiva sulle decisioni relative all'architettura e non un meccanismo di audit. Disporre di sistemi ben architettati aumenta notevolmente la probabilità di successo aziendale.

AWS Solutions Architects vanta anni di esperienza nella progettazione di soluzioni in un'ampia varietà di settori verticali aziendali e casi d'uso. Abbiamo supportato migliaia di clienti nella progettazione e revisione delle loro architetture su AWS. Grazie a questa esperienza, abbiamo identificato best practice e strategie principali per i sistemi di architettura nel cloud.

Il AWS Well-Architected Framework documenta una serie di domande fondamentali che aiutano a capire se un'architettura specifica si allinea bene alle best practice del cloud. Il framework fornisce un approccio coerente per la valutazione dei sistemi rispetto alle qualità che ti aspetti da sistemi basati sul cloud moderni e i rimedi necessari per raggiungere tali qualità. Man mano che la nostra evoluzione AWS continua e che continuiamo a imparare di più grazie alla collaborazione con i nostri clienti, continueremo a perfezionare la definizione di ben architettato.

Questo framework è destinato a coloro che ricoprono ruoli tecnologici, come i responsabili tecnologici (CTOs), gli architetti, gli sviluppatori e i membri del team operativo. Descrive le AWS migliori pratiche e strategie da utilizzare durante la progettazione e la gestione di un carico di lavoro cloud e fornisce collegamenti a ulteriori dettagli di implementazione e modelli architettonici. Per ulteriori informazioni, consulta la homepage di AWS Well-Architected.

AWS fornisce inoltre un servizio gratuito per la revisione dei carichi di lavoro. AWS Well-Architected Tool AWS (WA Tool) è un servizio nel cloud che fornisce un processo coerente per la revisione e la

misurazione dell'architettura utilizzando Well-Architected Framework. AWS Lo strumento AWS WA fornisce consigli per rendere i carichi di lavoro più affidabili, sicuri, efficienti ed economici.

Per aiutarti ad applicare le best practice, abbiamo creato AWS Well-Architected Labs, che fornisce un repository di codice e documentazione per un'esperienza concreta di implementazione delle best practice. Abbiamo anche collaborato con AWS partner selezionati di Partner Network (APN), che sono membri del programma AWS Well-Architected Partner. Questi AWS partner hanno una AWS conoscenza approfondita e possono aiutarti a rivedere e migliorare i tuoi carichi di lavoro.

Definizioni

Ogni giorno, gli esperti AWS assistono i clienti nell'architettura dei sistemi per sfruttare le migliori pratiche nel cloud. Ti aiutiamo a trovare i compromessi relativi all'architettura nel processo di evoluzione dei tuoi progetti. Quando implementi questi sistemi in ambienti live, analizziamo le prestazioni di questi sistemi e le conseguenze dei suddetti compromessi.

Sulla base di ciò che abbiamo appreso, abbiamo creato il AWS Well-Architected Framework, che fornisce a clienti e partner un insieme coerente di best practice per la valutazione delle architetture e fornisce una serie di domande che è possibile utilizzare per valutare l'allineamento di un'architettura alle best practice. AWS

Il AWS Well-Architected Framework si basa su sei pilastri: eccellenza operativa, sicurezza, affidabilità, efficienza delle prestazioni, ottimizzazione dei costi e sostenibilità.

Tabella 1. I pilastri del Framework AWS Well-Architected

Nome	Descrizione
Eccellenza operativa	Comprende la capacità di supportare lo sviluppo ed eseguire carichi di lavoro in modo efficace, ottenere informazioni approfondite sulle loro operazioni e migliorare continuam ente i processi e le procedure di supporto per offrire valore aggiunto.
Sicurezza	Il pilastro della sicurezza descrive come sfruttare le tecnologie cloud per proteggere dati, sistemi e risorse in modo da migliorare il livello di sicurezza.

Nome	Descrizione
Affidabilità	Il pilastro dell'affidabilità comprende la capacità di un carico di lavoro di eseguire la funzione attesa in modo corretto e coerente quando previsto. Ciò comprende la possibilità di utilizzare e testare il carico di lavoro per tutto il ciclo di vita. Questo paper fornisce una guida approfondita e sulle best practice per l'impleme ntazione di carichi di lavoro affidabili su. AWS
Efficienza delle prestazioni	La capacità di utilizzare in modo efficiente le risorse di elaborazione per soddisfare i requisiti di sistema e di mantenere tale efficienza di fronte al cambiamento delle richieste e all'evolu zione delle tecnologie.
Ottimizzazione dei costi	La capacità di eseguire i sistemi per distribuire il valore aziendale al prezzo minore.
Sostenibilità	La capacità di migliorare continuamente l'impatto sulla sostenibilità riducendo il consumo energetico e aumentando l'efficie nza di tutti i componenti di un carico di lavoro, massimizzando i benefici delle risorse allocate e riducendo al minimo le risorse totali richieste.

Nel AWS Well-Architected Framework, utilizziamo questi termini:

- Un componente è il codice, la configurazione e AWS le risorse che insieme soddisfano un requisito. Spesso un componente è l'unità di proprietà tecnica ed è disaccoppiato da altri componenti.
- Con il termine carico di lavoro ci riferiamo all'insieme di componenti che forniscono valore aziendale. Un carico di lavoro, normalmente, è il livello di dettaglio comunicato dai leader aziendali e della tecnologia.

- Secondo il nostro punto di vista, l'architettura è il modo in cui i componenti interagiscono in un carico di lavoro. Il modo di comunicare e di interagire dei componenti è spesso l'aspetto principale dei diagrammi architetturali.
- Le tappe fondamentali indicano cambiamenti chiave della tua architettura man mano che si evolve nel corso del ciclo di vita del prodotto (progettazione, test, messa online e produzione).
- Nell'ambito di un'organizzazione il portfolio delle tecnologie rappresenta l'insieme di carichi di lavoro necessari affinché l'azienda possa essere operativa.
- Il livello di impegno è la categorizzazione della quantità di tempo, sforzo e complessità che un'attività richiede per la sua realizzazione. Ogni organizzazione deve considerare le dimensioni e le competenze del team e la complessità del carico di lavoro per ottenere un contesto aggiuntivo che consenta di classificare correttamente il livello di impegno.
  - Elevato: il lavoro potrebbe richiedere più settimane o più mesi. Potrebbe essere suddiviso in molteplici fasi, rilasci e attività.
  - Medio: il lavoro potrebbe richiedere più giorni o settimane. Potrebbe essere suddiviso in molteplici rilasci e attività.
  - Basso: il lavoro potrebbe richiedere più ore o giorni. Potrebbe essere suddiviso in molteplici attività.

Quando progetti l'architettura dei carichi di lavoro, devi trovare dei compromessi tra i pilastri su cui si regge il tuo contesto aziendale. Le decisioni aziendali possono stabilire le priorità di progettazione. Potresti ottimizzare per migliorare la sostenibilità e ridurre i costi a spese dell'affidabilità in ambienti di sviluppo oppure, per quanto riguarda le soluzioni mission-critical, potresti ottimizzare l'affidabilità a fronte di costi più elevati e di un impatto ambientale maggiore. Nelle soluzioni di e-commerce, le prestazioni possono avere un impatto sui profitti e sulla propensione all'acquisto da parte dei clienti. Solitamente, la sicurezza e l'eccellenza operativa non sono soggette a compromessi rispetto agli altri pilastri.

Architettura

Negli ambienti on-premises, i clienti spesso hanno un team centrale per l'architettura delle tecnologie che funziona da livello superiore per altri team di prodotto o funzionalità, al fine di garantire che i team rispettino le best practice. I team dell'architettura delle tecnologie spesso sono composti da diversi ruoli come il Technical Architect (infrastruttura), il Solutions Architect (software), il Data Architect, il Networking Architect e il Security Architect. Spesso questi team utilizzano TOGAFo Zachman Framework come parte di una funzionalità di architettura aziendale.

In AWS, preferiamo distribuire le funzionalità tra i team piuttosto che avere un team centralizzato con tale capacità. Quando si sceglie di distribuire il potere decisionale si corrono dei rischi, ad esempio il rischio di garantire che i team interni rispettino gli standard. Noi mitighiamo questo rischi in due modi. In primo luogo, disponiamo di pratiche (modalità per eseguire attività, processi, standard e norme accettate) che hanno lo scopo di permettere a ogni team di possedere tali competenze e ci serviamo di esperti che verificano che i team adottino standard più severi di quelli che devono rispettare. In secondo luogo, implementiamo meccanismi che eseguono controlli automatizzati per verificare che gli standard vengano rispettati.

> "Le buone intenzioni non bastano mai, per avere successo servono buoni meccanismi", Jeff Bezos.

Questo significa sostituire gli sforzi di una persona con meccanismi (spesso automatizzati) che verificano la conformità alle regole e ai processi. Tale approccio distribuito è supportato dai principi di leadership di Amazon e stabilisce una cultura in tutti i ruoli che parte dal cliente. Il lavoro a ritroso è una parte fondamentale del nostro processo di innovazione. Partiamo dal cliente e da quello che vuole e sulla base di questo definiamo e indirizziamo i nostri sforzi. I team che mettono il cliente al centro sviluppano prodotti sulla base delle necessità del cliente.

Per l'architettura questo significa che ci aspettiamo che ogni team sia in grado di creare architetture e di seguire le best practice. Per aiutare i nuovi team ad acquisire queste capacità o i team esistenti ad alzare il livello, attiviamo l'accesso a una comunità virtuale di ingegneri principali che possono esaminare i loro progetti e aiutarli a capire quali sono le AWS migliori pratiche. La community di capo ingegneri lavora per rendere visibili e accessibili le best practice. Uno dei modi per fare ciò, ad esempio, è servirsi delle lunchtime talk che si concentrano sull'applicazione di best practice a esempi reali. Le lunchtime talk sono registrate e possono essere utilizzate come materiale di onboarding per i nuovi membri del team.

AWS le migliori pratiche emergono dalla nostra esperienza nella gestione di migliaia di sistemi su scala Internet. Preferiamo utilizzare i dati per definire le best practice, ma ci serviamo anche di esperti in materia, come i capo ingegneri. Quando i capo ingegneri vedono emergere nuove best practice, lavorano con la community per verificare che i team le rispettino. Con il tempo, queste best practice vengono formalizzate nei nostri processi di revisione interna e nei meccanismi che rafforzano la compliance. Il Framework Well-Architected è l'implementazione del nostro processo di revisione interno rivolta ai clienti, in cui abbiamo codificato la nostra idea di ingegneria responsabile attraverso

ruoli di campo come Solutions Architect e i team di ingegneria interni. Il Framework Well-Architected è un meccanismo scalabile che consente di trarre vantaggio da questi insegnamenti.

Seguendo l'approccio della community di capi ingegneri con la proprietà distribuita dell'architettura, riteniamo che si possa ottenere un'architettura aziendale Well-Architected che si basa sulle necessità del cliente. I leader tecnologici (come i responsabili dello sviluppo CTOs o dello sviluppo), effettuando revisioni di Well-Architected su tutti i carichi di lavoro, vi consentiranno di comprendere meglio i rischi del vostro portafoglio tecnologico. Tramite questo approccio puoi identificare dei temi tra i team che la tua organizzazione può affrontare tramite meccanismi, formazione o dialoghi informali in cui i capo ingegneri possono condividere le loro idee su aree specifiche con diversi team.

Principi generali di progettazione

Il Framework Well-Architected identifica una serie di principi generali per facilitare la corretta progettazione nel cloud:

- Smetti di ipotizzare quali siano le tue esigenze di capacità: se prendi una decisione sbagliata sulla capacità al momento dell'implementazione di un carico di lavoro, rischi di ritrovarti con risorse inattive o ad affrontare le conseguenze della capacità limitata. Con il cloud computing, questi problemi vengono risolti. Puoi utilizzare la capacità di cui hai bisogno e ridurre orizzontalmente o aumentare orizzontalmente il sistema automaticamente.
- Esegui test dei sistemi su scala produttiva: nel cloud, puoi creare un ambiente di test su scala produttiva on demand, completare i test e disattivare le risorse. Poiché paghi per l'ambiente di test solo quando è in esecuzione, puoi simulare un ambiente live a un costo notevolmente inferiore rispetti ai test on-premises.
- Automatizza pensando alla sperimentazione architettonica: l'automazione ti permette di creare e replicare i tuoi carichi di lavoro a basso costo e di evitare le spese della gestione manuale. Puoi tenere traccia delle modifiche all'automazione, effettuare l'audit dell'impatto e tornare ai parametri precedenti, se necessario.
- Considera le architetture evoluzionistiche: in un ambiente tradizionale, le decisioni relative all'architettura spesso sono implementate come eventi singoli e statici, con poche versioni principali di un sistema durante il ciclo di vita. Alla luce del continuo cambiamento di un'azienda e del suo contesto, le decisioni iniziali potrebbero ostacolare la capacità del sistema di soddisfare i requisiti aziendali in evoluzione. All'interno del cloud, la capacità di automatizzare e testare on demand diminuisce il rischio di impatto dovuto alle modifiche della progettazione. Questo permette ai sistemi di evolversi nel tempo, in modo che le aziende possano trarre vantaggio dalle innovazioni come pratica standard.

- Promuovi le architetture servendoti dei dati: nel cloud puoi raccogliere dati relativi all'impatto delle tue scelte architetturali sul comportamento del tuo carico di lavoro. Questo ti permette di prendere decisioni basate sui fatti su come migliorare il carico di lavoro. La tua infrastruttura cloud è un codice, quindi, puoi usare tali dati a vantaggio delle scelte e dei miglioramenti relativi all'architettura nel tempo.
- Migliora con le giornate di gioco: testa le prestazioni dell'architettura e dei processi pianificando regolarmente giornate di gioco per simulare eventi della produzione. Questi ti aiuta a capire dove puoi apportare dei miglioramenti e ti può aiutare a sviluppare un'esperienza organizzativa nella gestione degli eventi.

I pilastri del framework

La creazione di un sistema software è molto simile alla costruzione di un edificio. Se le fondamenta non sono solide, possono emergere problemi strutturali che minano l'integrità e la funzionalità dell'edificio. Se nella creazione dell'architettura per soluzioni tecnologiche trascuri i sei pilastri di eccellenza operativa, sicurezza, affidabilità, efficienza delle prestazioni, ottimizzazione dei costi e sostenibilità, può diventare complicato sviluppare un sistema che soddisfi le tue aspettative e i tuoi requisiti. L'aggiunta di questi pilastri alla tua architettura ti aiuterà a produrre sistemi efficienti e stabili. Questo ti permetterà di concentrarti su altri aspetti della progettazione, come i requisiti funzionali.

Pilastri

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

Eccellenza operativa

Il pilastro dell'eccellenza operativa comprende la capacità di supportare lo sviluppo ed eseguire carichi di lavoro in modo efficace, ottenere informazioni approfondite sulle loro operazioni e migliorare continuamente i processi e le procedure di supporto per offrire valore aggiunto.

Il pilastro dell'eccellenza operativa offre una panoramica dei principi di progettazione, delle best practice e delle domande. Puoi trovare linee guida prescrittive sull'implementazione nel whitepaper sul pilastro dell'eccellenza operativa.

Argomenti

- Principi di progettazione
- Definizione
- Best practice
- Risorse

Principi di progettazione

Ecco i principi di progettazione per l'eccellenza operativa nel cloud:

- Organizza i team in base ai risultati aziendali: la capacità di un team di conseguire i risultati aziendali deriva dalla visione della leadership, dall'efficacia delle operazioni e dall'allineamento del modello operativo all'azienda. La leadership dovrebbe essere pienamente investita e impegnata in una CloudOps trasformazione con un modello operativo cloud adeguato che incentivi i team a operare nel modo più efficiente e raggiungere i risultati di business. Il modello operativo corretto include persone, processi e capacità tecnologiche per scalare, ottimizzare la produttività e favorire la differenziazione tramite l'agilità, la reattività e l'adattamento. La visione a lungo termine dell'organizzazione si traduce in obiettivi che vengono comunicati alle parti interessate dell'azienda e agli utenti dei tuoi servizi cloud. Gli obiettivi e le attività operative KPIs sono allineati a tutti i livelli. Questa procedura promuove il valore a lungo termine derivante dall'implementazione dei seguenti principi di progettazione.
- Implementa l'osservabilità per approfondimenti utilizzabili: acquisisci una comprensione completa del comportamento, delle prestazioni, dell'affidabilità, dei costi e dello stato del carico di lavoro. Stabilisci gli indicatori chiave di performance (KPIs) e sfrutta la telemetria di osservabilità per prendere decisioni informate e agire tempestivamente quando i risultati aziendali sono a rischio. Migliora in modo proattivo le prestazioni, l'affidabilità e i costi sulla base di dati sull'osservabilità fruibili.
- Automatizza in modo sicuro, laddove possibile: nel cloud, ti è possibile applicare la medesima disciplina di progettazione che utilizzi per il codice dell'applicazione a tutto il tuo ambiente. Definisci l'intero carico di lavoro e le relative operazioni (applicazioni, infrastruttura, configurazione e procedure) come codice e aggiornarlo. Quindi, automatizza le operazioni del carico di lavoro avviandole in risposta agli eventi. Nel cloud, utilizzi la sicurezza dell'automazione configurando i guardrail, tra cui il controllo della frequenza, le soglie di errore e le approvazioni. Un'automazione efficiente offre risposte coerenti agli eventi, limita l'errore umano e riduce l'impegno degli operatori.
- Applica modifiche frequenti, minime e reversibili: progetta carichi di lavoro scalabili e con accoppiamento debole per consentire l'aggiornamento regolare dei componenti. Le tecniche di implementazione automatizzate insieme a modifiche incrementali più piccole riducono il raggio di esplosione, ovvero l'entità dell'impatto, e consentono un'inversione più rapida in caso di guasti. Ciò aumenta la fiducia necessaria per apportare modifiche strategiche al carico di lavoro mantenendo la qualità e adattandosi rapidamente ai cambiamenti delle condizioni di mercato.
- Perfeziona con frequenza le procedure operative: l'evoluzione delle operazioni deve seguire quella dei carichi di lavoro. Se usi procedure operative, cerca delle opportunità per migliorarle. Organizza

regolari revisioni per accertarti che tutte le procedure siano efficaci e che i team le conoscano adeguatamente. Se vengono individuate delle lacune, aggiorna le procedure di conseguenza. Comunica gli aggiornamenti procedurali a tutte le parti interessate e ai team. Converti le operazioni in gioco per condividere le best practice e fornire occasioni di formazione ai team.

- Prevedi gli insuccessi: massimizza il successo operativo definendo scenari di insuccesso per comprendere il profilo di rischio del carico di lavoro e il suo impatto sui risultati aziendali. Testa l'efficacia delle procedure e la risposta del team a questi errori simulati. Prendi decisioni informate per gestire i rischi aperti identificati tramite i test.
- Impara da tutti i parametri e gli eventi operativi: favorisci il miglioramento tramite le lezioni apprese da tutti gli eventi e gli errori operativi. Condividi ciò che hai imparato con i vari team e con tutta l'organizzazione. Gli insegnamenti evidenziano dati e aneddoti su come le operazioni contribuiscono al conseguimento dei risultati aziendali.
- Utilizza servizi gestiti: riduci l'onere operativo utilizzando i servizi gestiti, ove possibile. AWS Sviluppa procedure operative basate sulle interazioni con tali servizi.

Definizione

Esistono quattro aree di best practice per l'eccellenza operativa nel cloud:

- Organizzazione
- Preparazione
- Gestione
- Evoluzione

La leadership dell'organizzazione definisce gli obiettivi aziendali. La tua organizzazione deve comprendere i requisiti e le priorità e utilizzarli per organizzare e condurre attività a supporto del raggiungimento dei risultati aziendali. Il carico di lavoro deve generare le informazioni necessarie per supportarlo. L'implementazione di servizi per ottenere l'integrazione, l'implementazione e la distribuzione del carico di lavoro, darà vita a un flusso maggiore di modifiche vantaggiose in fase di produzione attraverso l'automazione dei processi ripetitivi.

Potrebbero esserci rischi inerenti al funzionamento del carico di lavoro. Occorre comprendere questi rischi e prendere una decisione consapevole prima di passare alla fase di produzione. I team devono essere in grado di supportare il carico di lavoro. Le metriche aziendali e operative derivate dai risultati aziendali desiderati ti aiuteranno a comprendere lo stato del carico di lavoro e le attività operative

e di rispondere agli incidenti. Le priorità cambieranno di pari passo con l'evoluzione delle esigenze aziendali e dell'ambiente aziendale. Utilizza questi aspetti come ciclo di feedback per apportare continui miglioramenti all'organizzazione e alle operazioni legate al carico di lavoro.

Best practice

> **Note**
>
> Tutte le domande sull'eccellenza operativa hanno il OPS prefisso come abbreviazione del pilastro.

Argomenti

Organizzazione

È necessario che i team abbiano una comprensione condivisa dell'intero carico di lavoro, del loro ruolo rispetto al carico di lavoro, nonché degli obiettivi aziendali condivisi. In questo modo potranno stabilire le priorità che possono favorire il successo aziendale. Un'adeguata definizione delle priorità massimizzerà i risultati dei tuoi sforzi. Valuta le esigenze dei clienti interni ed esterni coinvolgendo le principali parti interessate, compresi i team aziendali, di sviluppo e operativi, per stabilire dove concentrare le attività operative. Valutando le esigenze dei clienti otterrai una conoscenza approfondita del supporto necessario per raggiungere i risultati aziendali. Accertati di essere a conoscenza delle linee guida o degli obblighi definiti dalla governance organizzativa e da fattori esterni, come i requisiti di conformità normativa e gli standard di settore, che possono imporre o accentuare un'attenzione specifica. Accertati di disporre di meccanismi per identificare le modifiche ai requisiti di governance interna e di conformità esterni. Se non viene identificato alcun requisito, conferma l'applicazione della due diligence per giungere a tale determinazione. Rivedi regolarmente le tue priorità in modo che possano essere aggiornate al mutare delle esigenze.

Valuta le minacce per l'azienda (ad esempio rischi e responsabilità aziendali e minacce alla sicurezza delle informazioni) e conserva queste informazioni in un registro dei rischi. Valuta l'impatto dei rischi e dei compromessi tra interessi concorrenti o approcci alternativi. Ad esempio, accelerare l'introduzione

sul mercato di nuove funzionalità può essere preferibile all'ottimizzazione dei costi. Oppure, è possibile scegliere un database relazionale per i dati non relazionali per semplificare l'iniziativa di migrazione di un sistema senza rifattorizzare. Gestisci i vantaggi e i rischi per prendere decisioni informate nel determinare dove concentrare gli sforzi. Alcuni rischi o scelte possono essere accettabili per un certo periodo di tempo, potrebbe essere possibile ridurre i rischi associati o la presenza di un rischio potrebbe diventare inaccettabile, nel qual caso si intraprenderà un'azione per risolverlo.

I tuoi team devono comprendere quale contributo offrono nel raggiungimento dei risultati aziendali. I team devono avere obiettivi condivisi e comprendere il proprio ruolo nel successo degli altri team. Comprendere la responsabilità, la proprietà, il modo in cui vengono prese le decisioni e chi ha l'autorità decisionale aiuterà a concentrare gli sforzi e a ottimizzare i contributi dei team. Le esigenze di un team sono influenzate dal cliente supportato, dall'organizzazione, dalla composizione del team e dalle caratteristiche del carico di lavoro. Non è ragionevole aspettarsi che un singolo modello operativo sia in grado di supportare tutti i team e i relativi carichi di lavoro dell'organizzazione.

Assicurati che siano identificati i responsabili di ogni applicazione, carico di lavoro, piattaforma e componente dell'infrastruttura e che per ogni processo e procedura sia identificato un responsabile della definizione e dei responsabili delle prestazioni.

La comprensione del valore aziendale di ogni componente, processo e procedura, del motivo per cui tali risorse sono presenti o le attività vengono eseguite e del perché tale proprietà esiste indirizzerà le azioni dei membri del team. Definisci chiaramente le responsabilità dei membri del team in modo che possano agire in modo appropriato e disporre di meccanismi per identificare responsabilità e proprietà. Implementa meccanismi per richiedere aggiunte, modifiche ed eccezioni in modo da non porre limiti all'innovazione. Definisci gli accordi tra i team che descrivono il modo in cui collaborano per supportarsi reciprocamente e contribuire ai risultati aziendali.

Fornisci supporto ai membri del team in modo che possano essere più efficaci nell'azione e nel supporto dei risultati aziendali. La leadership aziendale di alto livello deve stabilire le aspettative e misurare il successo. La leadership aziendale di alto livello è promotrice, sostenitrice e motore per l'adozione delle best practice e l'evoluzione dell'organizzazione. Consenti ai membri del team di intervenire quando i risultati sono a rischio per ridurre al minimo l'impatto e incoraggiali a rivolgersi ai responsabili decisionali e alle parti interessate quando ritengono che esista un rischio, in modo da poterlo risolvere e prevenire gli incidenti. Fornisci comunicazioni tempestive, chiare e concrete dei rischi noti e degli eventi pianificati in modo che i membri del team possano agire in modo tempestivo e appropriato.

Incoraggia la sperimentazione per accelerare l'apprendimento e mantenere i membri del team interessati e coinvolti. I team devono aumentare le proprie competenze per adottare nuove

tecnologie e supportare i cambiamenti della domanda e delle responsabilità. Fornisci il tuo supporto e incoraggiamento offrendo tempo strutturato dedicato per l'apprendimento. Assicurati che i membri del team dispongano delle risorse, in termini sia di strumenti sia di membri del team, per avere successo e adattarsi, sostenendo i risultati aziendali. Sfrutta la diversità tra organizzazioni per cercare più prospettive uniche. Usa questa prospettiva per incrementare l'innovazione, mettere in discussione le tue ipotesi e ridurre il rischio di bias confermativi. Aumenta l'inclusione, la diversità e l'accessibilità all'interno dei team per ottenere prospettive vantaggiose.

Se esistono requisiti normativi e di conformità esterni applicabili alla tua organizzazione, utilizza le risorse fornite da AWS Cloud Compliance per promuovere la formazione dei tuoi team affinché siano in grado di valutare il relativo impatto sulle tue priorità. Il Framework Well-Architected enfatizza formazione, misurazione e miglioramento Fornisce un approccio coerente per valutare le architetture e implementare progetti scalabili nel tempo. AWS fornisce l'assistenza AWS Well-Architected Tool necessaria per rivedere l'approccio prima dello sviluppo, lo stato dei carichi di lavoro prima della produzione e lo stato dei carichi di lavoro in produzione. Puoi confrontare i carichi di lavoro con le migliori pratiche AWS architettoniche più recenti, monitorarne lo stato generale e ottenere informazioni sui potenziali rischi. AWS Trusted Advisor è uno strumento che fornisce l'accesso a una serie di controlli di base che consigliano ottimizzazioni che possono contribuire a definire le priorità. I clienti del supporto Business ed Enterprise hanno accesso a ulteriori controlli a livello di sicurezza, affidabilità, prestazioni, ottimizzazione dei costi e sostenibilità che possono essere utili per definire le loro priorità.

AWS può aiutarvi a istruire i vostri team in merito AWS ai relativi servizi per aumentare la loro comprensione di come le loro scelte possono avere un impatto sul vostro carico di lavoro. Utilizza le risorse fornite da AWS Support (AWS Knowledge Center, Forum di AWS discussione e AWS Support Centro) e la AWS documentazione per istruire i tuoi team. Rivolgiti AWS Support al AWS Support Centro per ricevere assistenza con le tue AWS domande. AWS condivide anche le migliori pratiche e i modelli che abbiamo appreso attraverso l'utilizzo di AWS The Amazon Builders' Library. Un'ampia varietà di altre informazioni utili è disponibile tramite il AWS blog e il podcast ufficiale AWS . AWS Training and Certification offre una certa formazione attraverso corsi digitali di autoapprendimento sui AWS fondamenti. Puoi anche iscriverti a un corso di formazione con istruttore per supportare ulteriormente lo sviluppo delle competenze dei tuoi team. AWS

Utilizza strumenti o servizi che ti consentano di governare centralmente i tuoi ambienti su più account, ad esempio per aiutarti a gestire i tuoi modelli operativi. AWS Organizations Servizi come AWS Control Tower ampliano questa capacità di gestione consentendovi di definire progetti (a supporto dei modelli operativi) per la configurazione degli account, applicare una governance continua utilizzando AWS Organizations e automatizzare il provisioning di nuovi account. I fornitori di servizi gestiti AWS

Managed Services, ad esempio AWS Managed Services Partner o Managed Services Providers del AWS Partner Network, offrono esperienza nell'implementazione di ambienti cloud e supportano i requisiti di sicurezza e conformità e gli obiettivi aziendali. L'aggiunta di servizi gestiti al tuo modello operativo ti consente di risparmiare tempo e risorse e ti permette di mantenere i team interni snelli e focalizzati sui risultati strategici che differenzieranno la tua attività, anziché sullo sviluppo di nuove competenze e funzionalità.

Le seguenti domande si concentrano su queste considerazioni relative all'eccellenza operativa. Per l'elenco completo delle domande e delle best practice relative all'eccellenza operativa, consulta l'Appendice.

OPS1: Come stabilisci quali sono le tue priorità?

È necessario che ognuno comprenda il proprio ruolo nel conseguimento del successo aziendale . Devi disporre di obiettivi comuni al fine di stabilire le priorità per le risorse. Ciò massimizzerà i risultati dei tuoi sforzi.

OPS2: Come strutturate la vostra organizzazione per supportare i risultati aziendali?

I tuoi team devono comprendere quale contributo offrono nel raggiungimento dei risultati aziendali. I team devono avere obiettivi condivisi e comprendere il proprio ruolo nel successo degli altri team. Comprendere la responsabilità, la proprietà, il modo in cui vengono prese le decisioni e chi ha l'autorità decisionale aiuterà a concentrare gli sforzi e a ottimizzare i contributi dei team.

OPS3: In che modo la vostra cultura organizzativa supporta i risultati aziendali?

Fornisci supporto ai membri del team in modo che possano essere più efficaci nell'azione e nel supporto dei risultati aziendali.

Ad esempio, a un certo punto potresti realizzare che desideri dare maggiore risalto a un piccolo sottoinsieme delle tue priorità. Utilizza un approccio equilibrato nel lungo termine per garantire lo sviluppo delle capacità necessarie e la gestione del rischio. Rivedi regolarmente le tue priorità e aggiornale al mutare delle esigenze. Quando la responsabilità e la proprietà sono indefinite o sconosciute, rischi sia di non affrontare tempestivamente le attività necessarie sia di adoperarti in

modo ridondante e potenzialmente conflittuale per rispondere a tali esigenze. La cultura organizzativa influisce direttamente sulla soddisfazione sul lavoro e sulla conservazione dei membri del team. Sostieni il coinvolgimento e le capacità dei membri del tuo team per ottenere il successo della tua attività. La sperimentazione è necessaria per realizzare l'innovazione e trasformare le idee in risultati. Un risultato indesiderato è un esperimento riuscito che ha identificato un percorso che non porterà al successo.

Preparazione

Per prepararti all'eccellenza operativa devi comprendere i carichi di lavoro e i loro comportamenti previsti. Sarai dunque in grado di progettare i carichi di lavoro in modo tale che forniscano informazioni sul loro stato e di creare le procedure per supportarli adeguatamente.

Progetta il tuo carico di lavoro affinché ti fornisca le informazioni necessarie a comprenderne lo stato interno (ad esempio, parametri, log, eventi e tracce) in tutti i componenti a supporto dell'osservabilità e dell'analisi dei problemi. L'osservabilità va oltre il semplice monitoraggio, in quanto fornisce una comprensione completa del funzionamento interno di un sistema basata sui suoi output esterni. L'osservabilità è legata a doppio filo a metriche, log e tracce per offrire informazioni approfondite sul comportamento e sulle dinamiche del sistema. Grazie a un'osservabilità efficace, i team possono distinguere modelli, anomalie e tendenze, così da essere in grado di affrontare in modo proattivo potenziali problemi e mantenere l'integrità del sistema. L'identificazione degli indicatori chiave di performance (KPIs) è fondamentale per garantire l'allineamento tra le attività di monitoraggio e gli obiettivi aziendali. Questo allineamento garantisce che i team prendano decisioni basate sui dati e su metriche realmente importanti, ottimizzando sia le prestazioni del sistema sia i risultati aziendali. Inoltre, l'osservabilità consente alle aziende di essere proattive anziché reattive. I team possono comprendere cause-and-effect le relazioni all'interno dei propri sistemi, prevedere e prevenire i problemi anziché limitarsi a reagire ad essi. Con l'evolversi dei carichi di lavoro, è essenziale riesaminare e perfezionare la strategia di osservabilità, assicurandosi che rimanga pertinente ed efficace.

Adotta strategie che migliorino il flusso delle modifiche in produzione e che consentano la rifattorizzazione, il feedback veloce sulla qualità e la correzione di errori. Tali prassi accelerano l'ingresso in produzione delle modifiche vantaggiose, limitano i problemi distribuiti e consentono una rapida identificazione e risoluzione dei problemi introdotti attraverso le attività di implementazione o scoperti negli ambienti.

Adotta prassi per fornire un feedback rapido sulla qualità e che permettano un ripristino veloce dalle modifiche che non hanno i risultati previsti. L'uso di queste prassi consente di mitigare l'impatto dei

problemi introdotti attraverso l'implementazione delle modifiche. Prepara un piano in caso di esito negativo delle modifiche in modo da poter rispondere più rapidamente se necessario, testando e convalidando le modifiche apportate. Sii consapevole delle attività pianificate nei tuoi ambienti in modo da poter gestire il rischio di modifiche che influiscono sulle attività pianificate. Privilegia le modifiche frequenti, piccole e reversibili per limitarne l'ambito. In questo modo velocizzerai risoluzione dei problemi e correzione, mantenendo la possibilità di rollback delle modifiche. In tal modo, è anche possibile ottenere più frequentemente i vantaggi offerti dalle modifiche importanti.

Valuta la prontezza operativa del carico di lavoro, dei processi e delle procedure, nonché del personale, per comprendere i rischi operativi correlati al carico di lavoro. Utilizza un processo omogeneo (inclusi elenchi di controllo manuali o automatici) per sapere quando puoi rilasciare un carico di lavoro o una modifica. Questo inoltre ti aiuterà a trovare le eventuali aree che necessitano di pianificazioni. Predisponi runbook che documentino le tue attività di routine e manuali alla base dei processi per la risoluzione dei problemi. Analizza i vantaggi e i rischi per prendere decisioni informate e consentire l'adozione delle modifiche nella produzione.

AWS consente di visualizzare l'intero carico di lavoro (applicazioni, infrastruttura, policy, governance e operazioni) come codice. In tal modo è possibile applicare la stessa disciplina ingegneristica utilizzata per il codice dell'applicazione a ogni elemento dello stack, condividendoli tra team o organizzazioni per sfruttare al massimo i vantaggi delle attività di sviluppo. Utilizza le operazioni come codice nel cloud e sfrutta la possibilità di sperimentare per sviluppare il tuo carico di lavoro e le procedure operative ed esercitarti con gli errori in modo sicuro. L'utilizzo AWS CloudFormation consente di disporre di ambienti di sviluppo, test e produzione coerenti e basati su modelli in modalità sandbox con livelli crescenti di controllo delle operazioni.

Le seguenti domande si concentrano su queste considerazioni relative all'eccellenza operativa.

OPS4: Come implementate l'osservabilità nel vostro carico di lavoro?

Implementare l'osservabilità nel carico di lavoro ti permette di comprendere lo stato di quest'ultimo e di adottare decisioni basate sui dati e che riflettono i requisiti aziendali.

OPS5: Come si riducono i difetti, si facilita la riparazione e si migliora il flusso di produzione?

Adotta strategie che migliorino il flusso delle modifiche in produzione e che favoriscano la rifattori zzazione, il feedback veloce sulla qualità e la correzione di errori. Tali approcci accelerano l'ingresso in produzione delle modifiche vantaggiose, contengono i problemi che si sono diffusi e

OPS5: Come si riducono i difetti, si facilita la riparazione e si migliora il flusso di produzione?

permettono di ottenere una rapida identificazione e risoluzione dei problemi introdotti attraverso le attività di implementazione.

OPS6: Come si mitigano i rischi di implementazione?

Adotta approcci per fornire un feedback rapido sulla qualità e che permettano un ripristino veloce dalle modifiche che non hanno i risultati previsti. L'uso di queste prassi consente di mitigare l'impatto dei problemi introdotti attraverso l'implementazione delle modifiche.

OPS7: Come fai a sapere di essere pronto a supportare un carico di lavoro?

Valuta la prontezza operativa del carico di lavoro, dei processi e delle procedure, nonché del personale per comprendere i rischi operativi correlati al carico di lavoro.

Investi nell'implementazione di attività operative come codice per aumentare al massimo la produttività del personale operativo, ridurre al minimo la frequenza degli errori e consentire risposte automatizzate. Utilizza l'analisi prefallimentare per prevedere errori e creare procedure ove opportuno. Applica i metadati utilizzando i Resource Tag e AWS Resource Groups seguendo una strategia di tagging coerente per identificare le tue risorse. Applica tag alle risorse per organizzare, monitorare i costi e controllare gli accessi e ottimizza l'esecuzione delle attività operative automatizzate. Adotta procedure di distribuzione che sfruttino l'elasticità del cloud per facilitare le attività di sviluppo e la pre-distribuzione dei sistemi e avere implementazioni più rapide. Quando apporti modifiche agli elenchi di controllo che utilizzi per valutare i tuoi carichi di lavoro, pianifica quello che farai con i sistemi live che non risultano più conformi.

Gestione

L'osservabilità ti consente di concentrarti su dati significativi e di comprendere le interazioni e l'output del tuo carico di lavoro. Concentrandoti sugli approfondimenti essenziali ed eliminando i dati non necessari, mantieni un approccio diretto alla comprensione delle prestazioni del carico di lavoro. È essenziale non solo raccogliere dati, ma anche interpretarli correttamente. Definisci linee guida chiare, imposta soglie di avviso appropriate e monitora attivamente eventuali deviazioni. Un cambiamento in una metrica chiave, specialmente se correlata ad altri dati, permette di individuare

aree problematiche specifiche. Grazie all'osservabilità hai strumenti per prevedere e affrontare potenziali sfide, assicurando che il tuo carico di lavoro funzioni senza intoppi e soddisfi le esigenze aziendali.

La corretta operatività di un carico di lavoro è misurata dal raggiungimento di risultati per l'azienda e per i clienti. Definisci i risultati desiderati, determina in che modo verrà misurato il successo e individua i parametri che saranno usati nei calcoli per determinare se il carico di lavoro e le operazioni sono efficaci. L'integrità delle operazioni include sia lo stato del carico di lavoro sia lo stato e il successo delle operazioni a supporto del carico di lavoro (ad esempio, l'implementazione e la risposta agli incidenti). Stabilisci le basi dei parametri per migliorare, eseguire indagini e intervenire, raccogliere e analizzare i parametri, quindi conferma la tua comprensione del successo operativo e della sua evoluzione nel corso del tempo. Usa i parametri raccolti per determinare il grado di soddisfazione dei clienti, capire se stai rispondendo alle esigenze aziendali e individuare gli aspetti da migliorare.

La gestione efficiente ed efficace degli eventi operativi è fondamentale per raggiungere l'eccellenza operativa. Ciò si applica agli eventi operativi sia pianificati che non. Usa runbook precisi per gli eventi chiari e ricorri ai playbook per favorire l'analisi e la risoluzione degli altri eventi. Attribuisci la priorità alle risposte agli eventi in base al loro impatto sull'azienda e sui clienti. Assicurati che, in caso di avvisi in risposta a un evento, vi sia una procedura associata da seguire, con un proprietario ben preciso. Definisci in anticipo il personale richiesto per risolvere un evento e includi dei processi di escalation per coinvolgere altro personale, ove necessario, in base all'urgenza e all'impatto. Individua e coinvolgi le persone che hanno l'autorità per prendere decisioni in merito alle linee d'azione laddove vi sia un impatto aziendale dovuto a una risposta a un evento non gestito precedentemente.

Comunica lo stato operativo dei carichi di lavoro tramite pannelli di controllo e notifiche personalizzati in base al pubblico di destinazione (ad esempio cliente, azienda, sviluppatori, addetti alle operazioni), in modo che gli interessati possano agire in maniera adeguata, che le loro aspettative vengano soddisfatte e che siano informati sulla ripresa delle normali operazioni.

In AWS, puoi generare visualizzazioni da dashboard delle metriche raccolte dai carichi di lavoro e in modo nativo da. AWS Puoi sfruttare CloudWatch le nostre applicazioni di terze parti per aggregare e presentare viste a livello aziendale, di carico di lavoro e operativo delle attività operative. AWS fornisce informazioni sul carico di lavoro attraverso funzionalità di registrazione AWS X-Ray, tra cui CloudWatch CloudTrail, e VPC Flow Logs per identificare i problemi del carico di lavoro a supporto dell'analisi e della correzione delle cause principali.

Le seguenti domande si concentrano su queste considerazioni relative all'eccellenza operativa.

OPS8: Come si utilizza l'osservabilità del carico di lavoro nella propria organizzazione?

Garantire l'integrità del carico di lavoro sfruttando l'osservabilità. Utilizzare metriche, log e tracce pertinenti per ottenere una visione completa delle prestazioni del carico di lavoro e risolvere i problemi in modo efficiente.

OPS9: Come comprendete lo stato di salute delle vostre operazioni?

Definisci, acquisisci e analizza i parametri delle operazioni per ottenere visibilità sugli eventi delle operazioni, in modo da intraprendere le azioni appropriate.

OPS10: Come gestite il carico di lavoro e gli eventi operativi?

Prepara e convalida le procedure in risposta agli eventi per ridurre al minimo il loro impatto sul tuo carico di lavoro.

Tutti i parametri raccolti devono essere allineati alle esigenze aziendali e ai risultati che supportano. Sviluppa risposte con script per eventi ben compresi e automatizza le prestazioni in risposta al riconoscimento dell'evento.

Evoluzione

Impara, condividi e migliora continuamente per sostenere l'eccellenza operativa. Dedica dei cicli di lavoro al raggiungimento di miglioramenti incrementali quasi continui. Esegui l'analisi post-incidente di tutti gli eventi che influiscono sul cliente. Identifica i fattori che contribuiscono e le azioni preventive per limitare o prevenire la ricorrenza. Comunica i fattori che contribuiscono alle comunità interessate, nel modo più adeguato. Valuta regolarmente e assegna le priorità alle opportunità di miglioramento (ad esempio, richieste di funzionalità, risoluzione dei problemi e requisiti di conformità), includendo sia il carico di lavoro sia le procedure operative.

Includi i loop di feedback nelle tue procedure per individuare rapidamente gli aspetti che devono essere migliorati e per acquisire conoscenze dall'esecuzione delle operazioni.

Condividi le lezioni apprese con i vari team per condividerne anche i vantaggi. Analizza le tendenze all'interno delle lezioni apprese ed esegui analisi trasversali retrospettive dei parametri operativi per

individuare le opportunità e i metodi di miglioramento. Implementa le modifiche previste per garantire il miglioramento e valuta i risultati per favorire il successo.

Sì AWS, puoi esportare i dati di log su Amazon S3 o inviare i log direttamente ad Amazon S3 per lo storage a lungo termine. Utilizzando AWS Glue, puoi scoprire e preparare i dati di log in Amazon S3 per l'analisi e archiviare i metadati associati in. AWS Glue Data Catalog Amazon Athena, grazie alla sua integrazione nativa con AWS Glue, può quindi essere utilizzato per analizzare i dati di log, interrogandoli utilizzando standard. SQL Utilizzando uno strumento di business intelligence come Amazon QuickSight, puoi visualizzare, esplorare e analizzare i tuoi dati. Rilevamento di tendenze ed eventi di interesse che possono portare a miglioramenti.

La seguente domanda si concentra su queste considerazioni relative all'eccellenza operativa.

OPS11: Come si evolvono le operazioni?
Dedica tempo e risorse per ottenere un miglioramento incrementale pressoché continuo, per far evolvere l'efficacia e l'efficienza delle tue operazioni.

L'evoluzione efficace delle operazioni si basa sugli elementi seguenti: miglioramenti piccoli ma frequenti; creazione di ambienti sicuri e tempo per sperimentare, sviluppare e testare i miglioramenti; ambienti in cui le persone siano incoraggiate a imparare dagli errori. Il supporto alle operazioni per ambienti sandbox, di sviluppo, di prova e di produzione, con un crescente livello di controlli operativi, facilita lo sviluppo e aumenta la prevedibilità dei risultati positivi dalle modifiche passate in produzione.

Risorse

Consulta le seguenti risorse per ulteriori informazioni sulle best practice relative all'eccellenza operativa.

Documentazione

- DevOps e AWS

Whitepaper

- Pilastro dell'eccellenza operativa

Video

- DevOps su Amazon

Sicurezza

Il pilastro della sicurezza contempla la capacità di proteggere dati, sistemi e asset per sfruttare le tecnologie cloud in modo da migliorare la sicurezza.

Il pilastro della sicurezza offre una panoramica dei principi di progettazione, delle best practice e delle domande. Puoi trovare linee guida prescrittive sull'implementazione nel whitepaper sul pilastro della sicurezza.

Argomenti

- Principi di progettazione
- Definizione
- Best practice
- Risorse

Principi di progettazione

Nel cloud sono presenti diversi principi utili per rafforzare la sicurezza del carico di lavoro:

- Implementa una solida base di identità: implementa il principio del privilegio minimo e applica la separazione dei compiti con l'autorizzazione appropriata per ogni interazione con le tue AWS risorse. Centralizza la gestione delle identità e mira a eliminare la dipendenza dalle credenziali statiche a lungo termine.
- Mantenimento della tracciabilità: monitora, crea avvisi e verifica in tempo reale le operazioni e le modifiche apportate al tuo ambiente. Integra la raccolta di log e parametri con i sistemi per analizzare e intervenire automaticamente.
- Applicazione della sicurezza a tutti i livelli: applica un approccio di difesa avanzata con più controlli di sicurezza. Applica a tutti i livelli (ad esempio, periferia della reteVPC, bilanciamento del carico, ogni istanza e servizio di elaborazione, sistema operativo, applicazione e codice).
- Automatizzazione delle best practice di sicurezza: i meccanismi di sicurezza automatizzati basati su software migliorano la capacità di scalare le risorse in modo sicuro, più rapido e conveniente.

Crea architetture sicure, compresa l'implementazione dei controlli, definite e gestite come codice nei modelli controllati dalle versioni.

- Protezione dei dati in transito e a riposo: classifica i dati in base a livelli di sensibilità e utilizza meccanismi quali crittografia, tokenizzazione e controllo degli accessi, ove opportuno.
- Accesso limitato delle persone ai dati: utilizza meccanismi e strumenti per ridurre o eliminare l'esigenza di accesso diretto o di elaborazione manuale dei dati. Ciò riduce il rischio di perdita, modifica e altri errori umani durante la gestione dei dati sensibili.
- Preparazione agli eventi di sicurezza: preparati per un incidente creando policy e processi di analisi e gestione degli incidenti in linea con i requisiti dell'organizzazione. Esegui simulazioni di risposta agli incidenti e utilizza strumenti dotati di automazione per aumentare la velocità nel rilevamento, nell'indagine e nel ripristino.

Definizione

Esistono sette aree di best practice per la sicurezza nel cloud.

- Nozioni di base sulla sicurezza
- Gestione dell'identità e degli accessi
- Rilevamento
- Protezione dell'infrastruttura
- Protezione dei dati
- Risposta agli incidenti
- Sicurezza delle applicazioni

Prima di progettare qualsiasi carico di lavoro, è necessario implementare pratiche che influenzano la sicurezza. Dovrai controllare chi può fare cosa. Inoltre, devi essere in grado di identificare gli incidenti di sicurezza, proteggere i tuoi sistemi e i tuoi servizi e mantenere la riservatezza e l'integrità dei dati attraverso la loro protezione. Dovresti avere dei processi ben definiti e rodati per rispondere a eventuali problemi di sicurezza. Questi strumenti e tecniche sono importanti perché supportano obiettivi come la prevenzione delle perdite finanziarie o la conformità agli obblighi normativi.

Il modello di responsabilità AWS condivisa aiuta le organizzazioni che adottano il cloud a raggiungere i propri obiettivi di sicurezza e conformità. Poiché protegge AWS fisicamente l'infrastruttura che supporta i nostri servizi cloud, come AWS cliente puoi concentrarti sull'utilizzo dei servizi per

raggiungere i tuoi obiettivi. Il AWS cloud offre inoltre un maggiore accesso ai dati di sicurezza e un approccio automatizzato alla risposta agli eventi di sicurezza.

Best practice

Argomenti

Sicurezza

La seguente domanda si concentra su queste considerazioni relative alla sicurezza. Per l'elenco completo delle domande e delle best practice relative alla sicurezza, consulta l'Appendice.

SEC1: Come gestite in modo sicuro il vostro carico di lavoro?

Per gestire il carico di lavoro in modo sicuro, è necessario applicare le best practice globali a ogni area di sicurezza. Segui i requisiti e i processi definiti in termini di eccellenza operativa a livello organizzativo e di carico di lavoro e applicali a tutte le aree.

Rimanere aggiornati con le raccomandazioni AWS, le fonti del settore e l'intelligence sulle minacce ti aiuta a far evolvere il tuo modello di minaccia e gli obiettivi di controllo. L'automazione dei processi di sicurezza, i test e la convalida permettono di dimensionare le operazioni di sicurezza.

In AWS, la separazione dei diversi carichi di lavoro per account, in base alla loro funzione e ai requisiti di conformità o sensibilità dei dati, è un approccio consigliato.

Gestione dell'identità e degli accessi

La gestione delle identità e degli accessi è una parte principale di un programma di sicurezza delle informazioni e garantisce che solo gli utenti e i componenti autorizzati e autenticati possano accedere

alle tue risorse e solo nella modalità che hai stabilito. Ad esempio, è necessario definire i principali (ovvero account, utenti, ruoli e servizi che possono eseguire operazioni nel tuo account), creare policy allineate a tali principali e implementare una forte gestione delle credenziali. Questi elementi a gestione privilegiata formano i concetti chiave dell'autenticazione e dell'autorizzazione.

Nel AWS, la gestione dei privilegi è supportata principalmente dal servizio AWS Identity and Access Management (IAM), che consente di controllare l'accesso utente e programmatico a AWS servizi e risorse. È necessario applicare criteri granulari che assegnano autorizzazioni a un utente, gruppo, ruolo o risorsa. È inoltre possibile richiedere procedure complesse in materia di password, ad esempio il livello di complessità, l'evitamento del riutilizzo e l'applicazione dell'autenticazione a più fattori (). MFA È possibile utilizzare la federazione con il servizio di directory esistente. Per i carichi di lavoro che richiedono l'accesso ai sistemi AWS, IAM consente un accesso sicuro tramite ruoli, profili di istanza, federazione delle identità e credenziali temporanee.

Le seguenti domande si concentrano su queste considerazioni relative alla sicurezza.

SEC2: Come si gestiscono le identità di persone e macchine?

Esistono due tipi di identità da gestire quando si tratta di gestire carichi di lavoro sicuri. AWS Comprendere il tipo di identità necessaria per gestire e concedere l'accesso ti aiuta a verificare che le identità corrette abbiano accesso alle risorse giuste nelle condizioni adeguate.

Identità umane: gli amministratori, gli sviluppatori, gli operatori e gli utenti finali richiedono un'identi tà per accedere agli ambienti e alle applicazioni. AWS Si tratta di membri dell'organizzazione o utenti esterni con cui collabora e che interagiscono con AWS le risorse dell'utente tramite un browser Web, un'applicazione client o strumenti interattivi da riga di comando.

Identità delle macchine: le applicazioni di servizio, gli strumenti operativi e i carichi di lavoro richiedono un'identità per effettuare richieste ai AWS servizi, ad esempio per leggere i dati. Queste identità includono macchine in esecuzione nel tuo AWS ambiente, come EC2 istanze o AWS Lambda funzioni Amazon. Puoi gestire le identità di macchine anche per soggetti esterni che necessitano dell'accesso. Inoltre, potresti avere anche macchine esterne AWS che richiedono l'accesso al tuo AWS ambiente.

SEC3: Come si gestiscono le autorizzazioni per persone e macchine?

Gestisci le autorizzazioni per controllare l'accesso alle identità di persone e macchine che richiedono l'accesso AWS e il tuo carico di lavoro. Le autorizzazioni controllano chi può accedere a cosa e a quali condizioni.

Le credenziali non devono essere condivise tra nessun utente o sistema. L'accesso degli utenti deve essere concesso utilizzando un approccio basato su privilegi minimi, basato sulle migliori pratiche, tra cui i requisiti in materia di password, e deve essere applicato. MFA L'accesso programmatico, comprese API le chiamate ai AWS servizi, deve essere eseguito utilizzando credenziali temporanee e con privilegi limitati, come quelle emesse da. AWS Security Token Service

Gli utenti necessitano di un accesso programmatico se desiderano interagire con l'esterno di. AWS AWS Management Console Il modo per concedere l'accesso programmatico dipende dal tipo di utente che accede. AWS

Per fornire agli utenti l'accesso programmatico, scegli una delle seguenti opzioni.

Quale utente necessita dell'accesso programmatico?	Per	Come
Identità della forza lavoro (Utenti gestiti in IAM Identity Center)	Utilizza credenziali temporane e per firmare le richieste programmatiche a AWS CLI, AWS SDKs, o. AWS APIs	Segui le istruzioni per l'interfa ccia che desideri utilizzare. • Per la AWS CLI, vedere Configurazione dell'uso AWS IAM Identity Center nella AWS CLI Guida per l'utente.AWS Command Line Interface • Per AWS SDKs gli strumenti e AWS APIs, consulta l'autenticazione di IAM Identity Center nella Guida di riferimento agli strumenti AWS SDKs e agli strumenti.

Quale utente necessita dell'accesso programmatico?	Per	Come
IAM	Utilizza credenziali temporane e per firmare le richieste programmatiche a AWS CLI, AWS SDKs, o. AWS APIs	Seguendo le istruzioni riportate in Utilizzo delle credenziali temporanee con le AWS risorse nella Guida per l'IAMutente.
IAM	(Non consigliato) Utilizza credenziali a lungo termine per firmare richieste programmatiche a AWS CLI,, AWS SDKs o. AWS APIs	Segui le istruzioni per l'interfa ccia che desideri utilizzare. • Per la AWS CLI, consulta Autenticazione tramite credenziali IAM utente nella Guida per l'utente.AWS Command Line Interface • Per AWS SDKs gli strumenti , consulta Autenticazione tramite credenziali a lungo termine nella Guida di riferimento agli strumenti e agli AWS SDKs strumenti. • Per AWS APIs, consulta Gestione delle chiavi di accesso per IAM gli utenti nella Guida per l'IAMutente.

AWS fornisce risorse che possono aiutarti nella gestione delle identità e degli accessi. Per imparare le best practice, scopri i nostri laboratori pratici sulla gestione delle credenziali e dell'autenticazione, sul controllo dell'accesso umano e sul controllo dell'accesso programmatico.

Rilevamento

Puoi utilizzare i controlli di rilevamento per identificare una potenziale minaccia o un potenziale incidente di sicurezza. Questi controlli sono una parte essenziale dei framework di governance e possono essere utilizzati per supportare il processo di qualità o un obbligo legale o di conformità

e per l'identificazione delle minacce e gli sforzi nelle risposte. Ci sono diversi tipi di controlli di rilevamento. Ad esempio, la realizzazione di un inventario di risorse e dei loro attributi dettagliati promuove le decisioni più efficienti (e i controlli del ciclo di vita) per stabilire delle baseline operative. Puoi anche utilizzare audit interni, un esame dei controlli relativi ai sistemi di informazioni, per verificare che le pratiche rispettino le policy e i requisiti e che tu abbia un set corretto di notifiche di avviso automatiche basate sulle condizioni definite. Questi controlli sono fattori di reazione importanti che possono aiutare la tua organizzazione a identificare e capire la portata dell'attività anomala.

Inoltre AWS, è possibile implementare controlli investigativi elaborando registri, eventi e monitoraggi che consentono il controllo, l'analisi automatizzata e l'invio di allarmi. CloudTrail registra, AWS API chiama e fornisce il monitoraggio delle metriche con allarmi e CloudWatch fornisce la cronologia delle configurazioni. AWS Config Amazon GuardDuty è un servizio gestito di rilevamento delle minacce che monitora continuamente i comportamenti dannosi o non autorizzati per aiutarti a proteggere i tuoi AWS account e i tuoi carichi di lavoro. Sono inoltre disponibili log a livello di servizio, ad esempio puoi utilizzare Amazon Simple Storage Service (Amazon S3) per registrare le richieste di accesso.

La seguente domanda si concentra su queste considerazioni relative alla sicurezza.

SEC4: Come rilevate e indagate sugli eventi di sicurezza?

Acquisisci e analizza gli eventi a partire da log e metriche per acquistare visibilità. Agisci su eventi di sicurezza e potenziali minacce per contribuire a rendere sicuro il carico di lavoro.

La gestione dei log è una parte importante di un carico di lavoro Well-Architected per ragioni che vanno da requisiti di sicurezza o forensi a disposizioni normative o legali. È fondamentale analizzare i log e rispondere in modo da identificare potenziali incidenti di sicurezza. AWS offre funzionalità che semplificano l'implementazione della gestione dei log, offrendo la possibilità di definire un ciclo di vita di conservazione dei dati o di definire dove verranno conservati, archiviati o eventualmente eliminati. Ciò rende la gestione dei dati prevedibile e affidabile, più semplice ed economica.

Protezione dell'infrastruttura

La protezione dell'infrastruttura comprende delle metodologie di controllo, come la difesa approfondita, necessarie per rispettare le best practice e gli obblighi organizzativi e normativi. L'utilizzo di queste metodologie è fondamentale per ottenere operazioni continuative e di successo sia nel cloud che on-premises.

In AWS, è possibile implementare l'ispezione dei pacchetti stateful e stateless, utilizzando tecnologie AWS native o utilizzando prodotti e servizi partner disponibili tramite. Marketplace AWSÈ necessario utilizzare Amazon Virtual Private Cloud (AmazonVPC) per creare un ambiente privato, sicuro e scalabile in cui definire la topologia, inclusi gateway, tabelle di routing e sottoreti pubbliche e private.

Le seguenti domande si concentrano su queste considerazioni relative alla sicurezza.

SEC5: Come proteggete le vostre risorse di rete?

Qualsiasi carico di lavoro che abbia una qualche forma di connettività di rete, che si tratti di Internet o di una rete privata, richiede più livelli di difesa per proteggere da minacce esterne e interne basate sulla rete.

SEC6: Come proteggete le vostre risorse di elaborazione?

Le risorse di calcolo nel carico di lavoro richiedono più livelli di difesa per contribuire alla protezion e da minacce esterne e interne. Le risorse di calcolo includono EC2 istanze, contenitori, AWS Lambda funzioni, servizi di database, dispositivi IoT e altro ancora.

Si consigliano più livelli di difesa in qualsiasi tipo di ambiente. Nel caso della protezione dell'infrastruttura, molti concetti e metodi sono validi sia per modelli cloud che on-premises. L'applicazione della protezione dei confini, il monitoraggio dei punti di ingresso e di uscita e la creazione di log, il monitoraggio e le notifiche completi sono tutti elementi essenziali per un efficace piano di sicurezza delle informazioni.

AWS i clienti possono personalizzare o rafforzare la configurazione di un contenitore Amazon Elastic Compute Cloud (Amazon)EC2, Amazon Elastic Container Service (AmazonECS) o di un' AWS Elastic Beanstalk istanza e mantenere tale configurazione su un'Amazon Machine Image immutabile (). AMI Quindi, indipendentemente dal fatto che vengano avviati tramite Auto Scaling o avviati manualmente, tutti i nuovi server virtuali (istanze) avviati con questo metodo AMI ricevono la configurazione avanzata.

Protezione dei dati

Prima di progettare qualsiasi sistema, devono essere stabiliti i requisiti fondamentali che influenzano la sicurezza. Ad esempio, la classificazione dei dati fornisce un modo per categorizzare i dati

organizzativi basati sui livelli di sensibilità, mentre la crittografia protegge i dati evitandone l'intelligibilità per gli accessi non autorizzati. Questi strumenti e tecniche sono importanti perché supportano obiettivi come la prevenzione delle perdite finanziarie o la conformità agli obblighi normativi.

Nel AWS, le seguenti pratiche facilitano la protezione dei dati:

- In qualità di AWS cliente, mantieni il pieno controllo sui tuoi dati.
- AWS semplifica la crittografia dei dati e la gestione delle chiavi, inclusa la rotazione regolare delle chiavi, che può essere facilmente automatizzata AWS o gestita dall'utente.
- È disponibile la creazione di log dettagliati con contenuti importanti, come l'accesso ai file e le modifiche.
- AWS ha progettato sistemi di storage per una resilienza eccezionale. Ad esempio, Amazon S3 Standard, S3 Standard-IA, One Zone-IA S3 e Amazon Glacier sono tutti progettati per offrire una resistenza degli oggetti del 99,999999999% in un determinato anno. Questo livello di durabilità corrisponde a una perdita media annua prevista dello 0,000000001% di oggetti.
- Il controllo delle versioni, che può far parte di un più ampio processo di gestione del ciclo di vita dei dati, può proteggere da sovrascritture accidentali, eliminazioni e danni simili.
- AWS non avvia mai lo spostamento dei dati tra le regioni. Il contenuto inserito in una regione rimarrà in quella regione a meno che tu non utilizzi esplicitamente una funzionalità o sfrutti un servizio che fornisce tale funzionalità.

Le seguenti domande si concentrano su queste considerazioni relative alla sicurezza.

SEC7: Come classificate i vostri dati?

La classificazione fornisce un modo per categorizzare i dati in base ai livelli di criticità e sensibilità, in modo da aiutarti a determinare i controlli di protezione e conservazione appropriati.

SEC8: Come proteggi i tuoi dati archiviati?

Proteggi i dati a riposo implementando più controlli per ridurre il rischio di accessi non autorizzati o altri comportamenti impropri.

SEC9: Come proteggete i vostri dati in transito?

Proteggi i dati in transito implementando più controlli per ridurre il rischio di accessi non autorizzati o perdita.

AWS offre diversi mezzi per crittografare i dati a riposo e in transito. Nei nostri servizi integriamo funzionalità che semplificano la crittografia dei dati. Ad esempio, abbiamo implementato la crittografia lato server (SSE) per Amazon S3 per semplificare l'archiviazione dei dati in forma crittografata. Puoi anche fare in modo che l'intero processo di HTTPS crittografia e decrittografia (generalmente noto come SSL terminazione) venga gestito da Elastic Load Balancing (). ELB

Risposta agli incidenti

Anche con controlli preventivi e investigativi estremamente maturi, la tua organizzazione dovrebbe comunque attuare processi per rispondere e mitigare il potenziale impatto di incidenti di sicurezza. L'architettura del carico di lavoro influisce fortemente sulla capacità dei team di operare efficacemente durante un incidente, isolare o contenere sistemi e ripristinare le operazioni a uno stato ottimale noto. La messa in atto degli strumenti e l'accesso prima di un incidente di sicurezza e la pratica sistematica della risposta agli incidenti durante i giorni di attività ti aiuterà a verificare che la tua architettura sia in grado di supportare indagini e ripristini tempestivi.

Nel AWS, le seguenti pratiche facilitano una risposta efficace agli incidenti:

- Sono disponibili log dettagliati che contengono contenuti importanti, come l'accesso ai file e le modifiche.
- Gli eventi possono essere elaborati automaticamente e avviare strumenti che automatizzano le risposte mediante l'uso di AWS APIs.
- Puoi effettuare il pre-provisioning degli strumenti e una "camera bianca" utilizzando AWS CloudFormation. In questo modo puoi effettuare indagini forensi in un ambiente sicuro e isolato.

La seguente domanda si concentra su queste considerazioni relative alla sicurezza.

SEC10: In che modo è possibile anticipare, reagire e riprendersi dagli incidenti?

La preparazione è cruciale per un esame tempestivo ed efficace degli incidenti di sicurezza, nonché per la risposta e il ripristino, così da ridurre al minimo potenziali interruzioni dell'orga nizzazione.

Verifica di poter garantire rapidamente l'accesso al tuo team addetto alla sicurezza e automatizzare l'isolamento delle istanze, oltre che acquisire i dati e lo stato per le indagini forensi.

Sicurezza delle applicazioni

La sicurezza delle applicazioni (AppSec) descrive il processo generale di progettazione, creazione e test delle proprietà di sicurezza dei carichi di lavoro sviluppati. Devi individuare persone sufficientemente qualificate nell'organizzazione, comprendere le proprietà di sicurezza dell'infrastruttura di sviluppo e rilascio e usare l'automazione per identificare i problemi correlati alla sicurezza.

L'adozione dei test di sicurezza delle applicazioni come parte regolare del ciclo di vita dello sviluppo del software (SDLC) e dei processi successivi al rilascio aiuta a convalidare l'esistenza di un meccanismo strutturato per identificare, correggere e prevenire i problemi di sicurezza delle applicazioni nell'ambiente di produzione.

La metodologia di sviluppo delle applicazioni deve includere controlli di sicurezza durante la progettazione, l'implementazione e il funzionamento dei carichi di lavoro. Nel frattempo, allinea il processo per una continua riduzione degli errori e l'azzeramento del debito tecnico. Ad esempio, usando la modellazione delle minacce durante la fase di progettazione, puoi individuare i difetti di progettazione e correggerli più facilmente e in modo meno costoso anziché attendere e mitigarli in un secondo momento.

Il costo e la complessità necessari per risolvere i difetti sono in genere inferiori quanto prima si entra nel. SDLC Il modo più semplice per risolvere i problemi è non averne affatto ed è per questo che un modello di rischio iniziale ti permette di concentrarti sui risultati corretti sin dalla fase di progettazione. Man mano che il AppSec programma matura, puoi aumentare la quantità di test eseguiti utilizzando l'automazione, migliorare la fedeltà del feedback ai costruttori e ridurre il tempo necessario per le revisioni di sicurezza. Tutte queste iniziative migliorano la qualità del software sviluppato e accelerano la distribuzione di funzionalità nell'ambiente di produzione.

Le presenti linee guida per l'implementazione si concentrano su quattro aree: organizzazione e cultura, sicurezza della pipeline, sicurezza nella pipeline e gestione delle dipendenze. Ogni area fornisce una serie di principi che è possibile implementare e fornisce una end-to-end panoramica di come progettare, sviluppare, creare, distribuire e gestire i carichi di lavoro.

Esistono diversi approcci che è possibile utilizzare per affrontare il programma di sicurezza delle applicazioni. AWS Alcuni sono basati sulla tecnologia, mentre altri sono incentrati sulle persone e gli aspetti organizzativi del programma.

La seguente domanda si concentra su queste considerazioni relative alla sicurezza dell'applicazione.

SEC11: Come si incorporano e si convalidano le proprietà di sicurezza delle applicazioni durante l'intero ciclo di vita di progettazione, sviluppo e distribuzione?

La formazione del personale, l'esecuzione di test tramite automazione, l'identificazione delle dipendenze e la convalida delle proprietà di sicurezza di strumenti e applicazioni riducono la probabilità del verificarsi di problemi di sicurezza nei carichi di lavoro di produzione.

Risorse

Consulta le seguenti risorse per ulteriori informazioni sulle best practice relative alla sicurezza.

Documentazione

- AWS Sicurezza nel cloud
- Conformità di AWS
- AWS Blog sulla sicurezza
- AWS Security Maturity Model

Whitepaper

- Pilastro della sicurezza
- AWS Panoramica sulla sicurezza
- AWS Rischio e conformità

Video

- AWS Stato di sicurezza dell'Unione
- Panoramica sulla responsabilità condivisa

Affidabilità

Il pilastro dell'affidabilità comprende la capacità di un carico di lavoro di eseguire la funzione attesa in modo corretto e coerente quando previsto. Ciò comprende la possibilità di utilizzare e testare il carico di lavoro per tutto il ciclo di vita. Questo paper fornisce una guida approfondita e sulle best practice per l'implementazione di carichi di lavoro affidabili su. AWS

Il pilastro dell'affidabilità offre una panoramica dei principi di progettazione, delle best practice e delle domande. Puoi trovare linee guida prescrittive sull'implementazione nel whitepaper sul pilastro dell'affidabilità.

Argomenti

- Principi di progettazione
- Definizione
- Best practice
- Risorse

Principi di progettazione

Esistono cinque principi di progettazione per l'affidabilità nel cloud:

- Ripristino automatico in caso di guasto: monitorando un carico di lavoro alla ricerca di indicatori chiave di prestazione (KPIs), è possibile avviare l'automazione quando viene superata una soglia. Questi KPIs dovrebbero essere una misura del valore aziendale, non degli aspetti tecnici del funzionamento del servizio. Ciò consente la notifica e il tracciamento automatici degli errori e i processi di recupero automatizzati che aggirano o riparano l'errore. Con un'automazione più sofisticata, è possibile anticipare e correggere gli errori prima che si verifichino.
- Test delle procedure di ripristino: in un ambiente on-premises, spesso vengono eseguiti test per dimostrare che il carico di lavoro funziona in uno scenario specifico. I test non vengono in genere utilizzati per convalidare le strategie di ripristino. Nel cloud, puoi testare il modo in cui il carico di lavoro incorre nell'errore e convalidare le procedure di ripristino. Puoi utilizzare l'automazione

per simulare diversi errori o ricreare scenari che in precedenza hanno portato a errori. Questo approccio presenta percorsi di errore che è possibile testare e correggere prima che si verifichi uno scenario di errore reale, riducendo così il rischio.

- Scalare a livello orizzontale per aumentare la disponibilità dei carichi di lavoro aggregati: sostituisci una risorsa grande con più risorse piccole per ridurre l'impatto di un singolo guasto sul carico di lavoro complessivo. Distribuisci le richieste tra più risorse di dimensioni inferiori per verificare che non condividano un punto di errore comune.
- Smetti di fare ipotesi sulla capacità: una causa comune di guasti nei carichi di lavoro on-premises è la saturazione delle risorse, quando le richieste assegnate a un carico di lavoro superano la capacità di quel carico di lavoro (questo è spesso l'obiettivo di attacchi di tipo Denial of Service). Nel cloud, è possibile monitorare la domanda e l'utilizzo dei carichi di lavoro, nonché automatizzare l'aggiunta o la rimozione di risorse per mantenere il livello più efficiente, al fine di soddisfare la domanda senza un provisioning eccessivo o inferiore. Esistono ancora limiti, ma alcune quote possono essere controllate e altre possono essere gestite (consulta Gestione di Service Quotas e vincoli).
- Gestione del cambiamento tramite l''automazione: le modifiche all'infrastruttura andrebbero apportate utilizzando l'automazione. Le modifiche che devono essere gestite includono quelle all'automazione, che possono quindi essere monitorate e revisionate.

Definizione

Esistono quattro aree di best practice per l'affidabilità nel cloud:

- Fondamenti
- Architettura del carico di lavoro
- Gestione delle modifiche
- Gestione dei guasti

Per ottenere affidabilità, è necessario iniziare dalle basi: un ambiente in cui Service Quotas e topologia di rete sono in grado di supportare il carico di lavoro. L'architettura del carico di lavoro del sistema distribuito deve essere progettata per prevenire e mitigare i guasti. Il carico di lavoro deve gestire le variazioni nella domanda o nei requisiti e deve essere progettato per rilevare il guasto e applicare autonomamente le correzioni in automatico.

Best practice

Argomenti

Fondamenti

I requisiti di base sono quelli il cui ambito si estende oltre un singolo carico di lavoro o progetto. Prima di progettare qualsiasi sistema, occorre stabilire i requisiti fondamentali che influenzano l'affidabilità. Ad esempio, è necessario disporre di una larghezza di banda della rete sufficiente verso il data center.

Attualmente AWS, la maggior parte di questi requisiti fondamentali sono già incorporati o possono essere soddisfatti secondo necessità. Il cloud è progettato per essere pressoché illimitato, quindi è responsabilità di AWS soddisfare il requisito di una capacità di rete e di elaborazione sufficiente, che consenta di modificare le dimensioni e le allocazioni delle risorse su richiesta.

Le seguenti domande si concentrano su queste considerazioni relative all'affidabilità. Per l'elenco completo delle domande e delle best practice relative all'affidabilità, consulta l'Appendice.

REL1: Come gestite i Service Quotas e i vincoli?

Per le architetture di carichi di lavoro basate sul cloud, esistono Service Quotas (chiamate anche restrizioni dei servizi). Queste quote servono a prevenire l'approvvigionamento accidentale di più risorse del necessario e a limitare la frequenza delle richieste relative alle API operazioni in modo da proteggere i servizi da eventuali abusi. Esistono anche vincoli di risorse, ad esempio la velocità con cui è possibile trasferire i bit su un cavo in fibra ottica o lo spazio di archiviazione su un disco fisico.

REL2: Come si pianifica la topologia di rete?

I carichi di lavoro sono spesso presenti in più ambienti. Questi includono più ambienti cloud (sia accessibili pubblicamente sia privati) e, possibilmente, l'infrastruttura del data center esistente. I

REL2: Come si pianifica la topologia di rete?

piani devono includere considerazioni di rete, ad esempio connettività intrasistema e intersistema, gestione di indirizzi IP pubblici, gestione di indirizzi IP privati e risoluzione dei nomi di dominio.

Architettura del carico di lavoro

Un carico di lavoro affidabile comincia con decisioni iniziali di progettazione sia per il software sia per l'infrastruttura. Le tue scelte architetturali avranno un impatto sul comportamento del carico di lavoro su tutti i pilastri del Framework Well-Architected. Per l'affidabilità, è necessario seguire modelli specifici.

Con AWS, gli sviluppatori di carichi di lavoro possono scegliere i linguaggi e le tecnologie da utilizzare. AWS SDKselimina la complessità della programmazione fornendo servizi specifici per ogni linguaAPIs. AWS QuestiSDKs, oltre alla scelta dei linguaggi, consentono agli sviluppatori di implementare le migliori pratiche di affidabilità elencate qui. Gli sviluppatori possono anche leggere e scoprire come Amazon crea e gestisce software nella Amazon Builders' Library.

Le seguenti domande si concentrano su queste considerazioni relative all'affidabilità.

REL3: Come si progetta l'architettura dei servizi per i carichi di lavoro?

Crea carichi di lavoro altamente scalabili e affidabili utilizzando un'architettura orientata ai servizi (SOA) o un'architettura di microservizi. L'architettura orientata ai servizi (SOA) è la pratica di rendere i componenti software riutilizzabili tramite interfacce di servizio. L'architettura dei microservizi va oltre, per rendere i componenti più piccoli e semplici.

REL4: Come si progettano le interazioni in un sistema distribuito per prevenire i guasti?

I sistemi distribuiti si basano sulle reti di comunicazione per interconnettere i componenti, ad esempio server o servizi. Il carico di lavoro deve funzionare in modo affidabile nonostante la perdita o la latenza dei dati su queste reti. I componenti del sistema distribuito devono funzionar e in modo da non influire negativamente su altri componenti o sul carico di lavoro. Queste best practice prevengono i guasti e migliorano il tempo medio tra i guasti (). MTBF

REL5: Come si progettano le interazioni in un sistema distribuito per mitigare o resistere ai guasti?

I sistemi distribuiti si basano sulle reti di comunicazione per interconnettere i componenti (ad esempio server o servizi). Il carico di lavoro deve funzionare in modo affidabile nonostante la perdita o la latenza dei dati su queste reti. I componenti del sistema distribuito devono funzionar e in modo da non influire negativamente su altri componenti o sul carico di lavoro. Queste best practice permettono ai carichi di lavoro di tollerare le sollecitazioni o i guasti, recuperare più rapidamente e mitigare l'impatto di tali problemi. Il risultato è un miglioramento del tempo medio di ripristino ()MTTR.

Gestione delle modifiche

Le modifiche apportate al carico di lavoro o al relativo ambiente devono essere previste e gestite in modo da garantire l'affidabilità del carico di lavoro. Certe modifiche al carico di lavoro sono imposte da fattori esterni, quali i picchi di domanda, e anche altre modifiche dipendono da fattori interni, quali le distribuzioni delle funzionalità e le patch di sicurezza.

Utilizzando AWS, è possibile monitorare il comportamento di un carico di lavoro e automatizzare la risposta a. KPIs Ad esempio, il carico di lavoro può aggiungere ulteriori server man mano che il carico di lavoro acquisisce più utenti. È possibile controllare chi dispone dell'autorizzazione per apportare modifiche al carico di lavoro ed eseguire l'audit della cronologia di tali modifiche.

Le seguenti domande si concentrano su queste considerazioni relative all'affidabilità.

REL6: Come si monitorano le risorse del carico di lavoro?

I log e le metriche sono strumenti molto efficaci per ottenere informazioni sullo stato del carico di lavoro. Puoi configurare il carico di lavoro in modo da monitorare i log e le metriche e inviare notifiche in caso di superamento delle soglie o di eventi significativi. Il monitoraggio permette al carico di lavoro di riconoscere il superamento delle soglie di prestazioni basse o il verificarsi di errori, in modo da ripristinarlo in automatico di conseguenza.

REL7: Come si progetta il carico di lavoro per adattarlo ai cambiamenti della domanda?

Un carico di lavoro scalabile garantisce l'elasticità per aggiungere o rimuovere risorse in automatic o, in modo che sussista una stretta corrispondenza con la domanda attuale in un dato momento.

REL8: Come implementate il cambiamento?

Per implementare nuove funzionalità e verificare che i carichi di lavoro e l'ambiente operativo eseguano software noti e che sia possibile applicare patch o sostituirli in modo prevedibile, sono necessarie modifiche controllate. Se invece non sono controllate, risulta difficile prevederne l'effetto o risolvere eventuali problemi che causano.

Progettando un carico di lavoro in grado di aggiungere e rimuovere automaticamente le risorse in risposta ai cambiamenti della domanda, non solo si aumenta l'affidabilità, ma si convalida anche che il successo aziendale non diventi un peso. Con il monitoraggio in atto, il tuo team verrà avvisato automaticamente in caso di KPIs deviazione dalle norme previste. La registrazione automatica delle modifiche al proprio ambiente permette di controllare e identificare rapidamente le azioni che potrebbero avere influito sull'affidabilità. I controlli sulla gestione delle modifiche certificano la possibilità di applicare le regole che garantiscono l'affidabilità di cui hai bisogno.

Gestione dei guasti

In qualsiasi sistema di ragionevole complessità è previsto che si verifichino errori. L'affidabilità richiede che il carico di lavoro venga a conoscenza degli errori nel momento in cui si verificano e intervenga per evitare conseguenze sulla disponibilità. I carichi di lavoro devono essere in grado di affrontare errori e risolvere automaticamente i problemi.

Con AWS, puoi sfruttare l'automazione per reagire ai dati di monitoraggio. Ad esempio, quando un determinato parametro supera una soglia, è possibile avviare un'azione automatizzata per risolvere il problema. Inoltre, anziché tentare di diagnosticare e correggere una risorsa guasta che fa parte del tuo ambiente di produzione, puoi sostituirla con una nuova ed eseguire l'analisi sulla risorsa guasta fuori banda. Poiché il cloud consente di creare versioni temporanee di un intero sistema a basso costo, è possibile utilizzare i test automatizzati per verificare i processi di recupero completi.

Le seguenti domande si concentrano su queste considerazioni relative all'affidabilità.

REL9: Come si esegue il backup dei dati?

Esegui il backup di dati, applicazioni e configurazione per soddisfare i requisiti relativi agli obiettivi dei tempi di ripristino (RTO) e agli obiettivi dei punti di ripristino (RPO).

REL10: Come si utilizza l'isolamento dei guasti per proteggere il carico di lavoro?

Le barriere per l'isolamento dei guasti limitano l'effetto di un errore all'interno di un carico di lavoro a un numero limitato di componenti. I componenti al di fuori della barriera non subiscono gli effetti del guasto. Utilizzando più barriere per l'isolamento dei guasti, puoi limitare l'impatto sul carico di lavoro.

REL11: Come si progetta il carico di lavoro per resistere ai guasti dei componenti?

I carichi di lavoro che richiedono un'elevata disponibilità e un basso tempo medio di ripristino (MTTR) devono essere progettati per garantire la resilienza.

REL12: Come si verifica l'affidabilità?

Dopo aver progettato il carico di lavoro in modo da essere resiliente alle sollecitazioni della produzione, i test sono l'unico modo per verificare il funzionamento corretto e offrire la resilienza prevista.

REL13: Come si pianifica il disaster recovery (DR)?

Avere backup e componenti del carico di lavoro ridondanti in loco è l'inizio della strategia di ripristino di emergenza. RTOe RPO sono i vostri obiettivi per il ripristino del carico di lavoro. Imposta questi valori in base alle esigenze aziendali. Implementa una strategia per raggiungere questi obiettivi, prendendo in considerazione le posizioni e la funzione delle risorse e dei dati del carico di lavoro. La probabilità di interruzione e il costo del ripristino sono fattori chiave che aiutano a comunicare il valore aziendale che può avere il ripristino di emergenza per un carico di lavoro.

Esegui regolarmente il backup dei dati e testa i file di backup per verificare di poter effettuare il ripristino dopo errori sia logici che fisici. Una chiave per la gestione dei guasti è il test frequente e automatico dei carichi di lavoro che causano gli errori e quindi osservare come si ripristinano. Esegui questa operazione regolarmente e verifica che tali test vengano avviati anche dopo importanti cambiamenti del carico di lavoro. Monitora attivamenteKPIs, oltre all'obiettivo del tempo di ripristino (RTO) e all'obiettivo del punto di ripristino (RPO), per valutare la resilienza di un carico di lavoro

(specialmente in scenari di failure-test). Il monitoraggio KPIs ti aiuterà a identificare e mitigare i singoli punti di errore. L'obiettivo è testare a fondo i processi di ripristino del carico di lavoro in modo da avere la certezza di poter recuperare tutti i dati e continuare a servire i propri clienti, anche di fronte a problemi prolungati. I processi di recupero dovrebbero essere testati tanto quanto i normali processi di produzione.

Risorse

Consulta le seguenti risorse per ulteriori informazioni sulle best practice per l'affidabilità.

Documentazione

- Documentazione AWS
- AWS Infrastruttura globale
- AWS Auto Scaling: How Scaling Plans Work
- Che cos'è AWS Backup?

Whitepaper

- Pilastro dell'affidabilità: AWS Well-Architected
- Implementazione di microservizi su AWS

Efficienza delle prestazioni

Il pilastro dell'efficienza delle prestazioni include la capacità di utilizzare in modo efficiente le risorse nel cloud per soddisfare i requisiti in termini di prestazione e di mantenere tale efficienza a fronte al cambiamento della domanda e all'evoluzione delle tecnologie.

Il pilastro dell'efficienza delle prestazioni offre una panoramica dei principi di progettazione, delle best practice e delle domande. Puoi trovare linee guida prescrittive sull'implementazione nel whitepaper sul pilastro dell'efficienza delle prestazioni.

Argomenti

- Principi di progettazione
- Definizione
- Best practice

- Risorse

Principi di progettazione

Esistono cinque principi di progettazione per l'efficienza delle prestazioni nel cloud:

- Estendi a tutti le tecnologie avanzate: agevola l'implementazione di tecnologie avanzate da parte del tuo team delegando le attività complesse al tuo fornitore di cloud. Anziché chiedere al team IT di imparare come adottare e gestire una nuova tecnologia, valuta l'opportunità di utilizzare la tecnologia come servizio. Ad esempio, No SQL database, transcodifica multimediale e apprendimento automatico sono tutte tecnologie che richiedono competenze specialistiche. Nel cloud, tali tecnologie diventano servizi che il tuo team può semplicemente utilizzare mentre si concentra sullo sviluppo di un prodotto invece che sul provisioning e sulla gestione delle risorse.
- Diventa globale in pochi minuti: l'implementazione del carico di lavoro in più AWS regioni del mondo ti consente di fornire una latenza inferiore e un'esperienza migliore ai tuoi clienti a costi minimi.
- Utilizza architetture serverless: scegliendo le architetture serverless, non avrai più bisogno di gestire e mantenere server fisici per portare a termine le attività di elaborazione tradizionali. Ad esempio, i servizi di storage serverless possono agire da siti web statici, eliminando la necessità di server web, mentre i servizi di eventi possono ospitare il codice. Questo elimina l'onere operativo della gestione dei server fisici, con una riduzione dei costi delle transazioni, dal momento che servizi gestiti di questo tipo funzionano a livello di cloud.
- Sperimenta più di frequente: le risorse virtuali e automatizzabili ti permettono di portare a termine velocemente i test comparativi utilizzando diversi tipi di istanze, storage o configurazioni.
- Prendi in considerazione la comprensione meccanica: scopri come vengono consumati i servizi cloud e utilizza sempre l'approccio tecnologico più adatto ai tuoi obiettivi di carico di lavoro. Ad esempio, prendi in considerazione gli schemi di accesso ai dati quando selezioni una strategia basata su database o archiviazione.

Definizione

Esistono cinque aree di best practice per l'efficienza delle prestazioni nel cloud:

- Scelta dell'architettura
- Calcolo e hardware
- Gestione dei dati

- Reti e distribuzione di contenuti
- Processo e cultura

Utilizza un approccio basato sui dati per la creazione di un'architettura a prestazioni elevate. Raccogli dati su tutti gli aspetti dell'architettura, dalla progettazione di alto livello alla selezione e alla configurazione dei tipi di risorse.

La revisione periodica delle tue scelte conferma che stai sfruttando il cloud in continua evoluzione. AWS Il monitoraggio ti assicurerà di essere consapevole di qualsiasi divergenza rispetto alle prestazioni previste. Infine, puoi raggiungere dei compromessi nella tua architettura per migliorare le prestazioni, ad esempio utilizzando la compressione o la memorizzazione nella cache oppure allentando i requisiti di coerenza.

Best practice

Argomenti

Scelta dell'architettura

La soluzione ottimale per un determinato carico di lavoro può variare e le soluzioni spesso combinano molteplici approcci. I carichi di lavoro Well-Architected utilizzano soluzioni multiple e forniscono funzionalità diverse per migliorare le prestazioni.

AWS le risorse sono disponibili in molti tipi e configurazioni, il che rende più facile trovare un approccio che corrisponda strettamente alle proprie esigenze. Inoltre, puoi trovare opzioni che non sono facili da trovare nelle infrastrutture on-premises. Ad esempio, un servizio gestito come Amazon DynamoDB fornisce un database SQL No completamente gestito con latenza di un millisecondo su qualsiasi scala.

Le seguenti domande si concentrano su queste considerazioni relative all'efficienza delle prestazioni. Per l'elenco completo delle domande e delle best practice relative all'efficienza delle prestazioni, consulta l'Appendice.

PERF1: Come selezionate le risorse cloud e i modelli di architettura appropriati per il vostro carico di lavoro?

Spesso sono necessari molteplici approcci per ottenere prestazioni più efficienti in un carico di lavoro. I sistemi Well-Architected utilizzano più soluzioni e funzionalità per migliorare le prestazio ni.

Calcolo e hardware

La soluzione ottimale in termini di calcolo per un determinato carico di lavoro potrebbe variare in base alla progettazione dell'applicazione, ai modelli di utilizzo e alle impostazioni di configurazione. Le architetture possono utilizzare diverse soluzioni di calcolo per vari componenti e impiegare funzionalità diverse per migliorare le prestazioni. Selezionare la soluzione di calcolo sbagliata per un'architettura può ridurre l'efficienza delle prestazioni.

In AWS, il calcolo è disponibile in tre forme: istanze, contenitori e funzioni:

- Le istanze sono server virtualizzati che consentono di modificarne le funzionalità con un pulsante o una chiamata. API Poiché nel cloud le decisioni relative alle risorse non sono cristallizzate nel tempo, è possibile sperimentare vari tipi di server. Attualmente AWS, queste istanze di server virtuali sono disponibili in famiglie e dimensioni diverse e offrono un'ampia varietà di funzionalità, tra cui unità a stato solido (SSDs) e unità di elaborazione grafica (). GPUs
- I container sono un metodo di virtualizzazione del sistema operativo che consente di eseguire un'applicazione e le sue dipendenze in processi con risorse isolate. AWS Fargate è un'elaborazione serverless per contenitori oppure Amazon EC2 può essere utilizzato se hai bisogno di controllare l'installazione, la configurazione e la gestione del tuo ambiente di elaborazione. Puoi anche scegliere tra più piattaforme di orchestrazione dei container: Amazon Elastic Container Service (ECS) o Amazon Elastic Kubernetes Service (). EKS
- Le funzioni astraggono l'ambiente di esecuzione dal codice che desideri eseguire. Ad esempio, AWS Lambda consente di eseguire codice senza eseguire un'istanza.

Le seguenti domande si concentrano su queste considerazioni relative all'efficienza delle prestazioni.

PERF2: Come selezionate e utilizzate le risorse di calcolo nel vostro carico di lavoro?

La soluzione di calcolo più efficiente per un determinato carico di lavoro varia in base alla progettazione dell'applicazione, ai modelli di utilizzo e alle impostazioni di configurazione. Le architetture possono utilizzare diverse soluzioni di elaborazione per vari componenti e attivare funzionalità diverse per migliorare le prestazioni. Selezionare la soluzione di calcolo sbagliata per un'architettura può portare a una riduzione dell'efficienza delle prestazioni.

Gestione dei dati

La soluzione di gestione dei dati ottimale per un particolare sistema varia in base al tipo di dati (blocco, file o oggetto), ai modelli di accesso (casuale o sequenziale), alla velocità effettiva richiesta, alla frequenza di accesso (online, offline, di archiviazione), alla frequenza di aggiornamento (WORMdinamica) e ai vincoli di disponibilità e durabilità. I carichi di lavoro Well-Architected utilizzano archivi dati appositamente progettati che impiegano diverse funzionalità per migliorare le prestazioni.

In AWS, lo storage è disponibile in tre forme: oggetto, blocco e file:

- Lo storage a oggetti fornisce una piattaforma scalabile e durevole per rendere i dati accessibili da qualsiasi posizione Internet per contenuti generati dagli utenti, archivi attivi, computing serverless, storage di big data o backup e ripristino. Amazon Simple Storage Service (Amazon S3) è un servizio di archiviazione di oggetti che offre scalabilità, disponibilità dei dati, sicurezza e prestazioni tra le migliori del settore. Amazon S3 è progettato per garantire una durabilità del 99,999999999% (11 9) e memorizza i dati per milioni di applicazioni per aziende in tutto il mondo.
- Lo storage a blocchi offre uno storage a blocchi ad alta disponibilità, coerente e a bassa latenza per ogni host virtuale ed è analogo allo storage collegato direttamente (DAS) o allo Storage Area Network (). SAN Amazon Elastic Block Store (AmazonEBS) è progettato per carichi di lavoro che richiedono uno storage persistente accessibile da EC2 istanze che consente di ottimizzare le applicazioni con la capacità di storage, le prestazioni e i costi corretti.
- Lo storage di file fornisce accesso a un file system condiviso tra più sistemi. Le soluzioni di storage di file come Amazon Elastic File System (AmazonEFS) sono ideali per casi d'uso come archivi di contenuti di grandi dimensioni, ambienti di sviluppo, negozi multimediali o home directory degli utenti. Amazon FSx rende efficiente ed economico il lancio e l'esecuzione dei file system più diffusi in modo da poter sfruttare i ricchi set di funzionalità e le prestazioni rapide dei file system open source e con licenza commerciale ampiamente utilizzati.

Le seguenti domande si concentrano su queste considerazioni relative all'efficienza delle prestazioni.

PERF3: Come archiviate, gestite e accedete ai dati del vostro carico di lavoro?

La soluzione di storage più efficiente per un sistema varia in base al tipo di operazione di accesso (blocco, file o oggetto), ai modelli di accesso (casuale o sequenziale), al throughput richiesto , alla frequenza di accesso (online, offline, di archiviazione), alla frequenza di aggiornamento (WORMdinamica) e ai vincoli di disponibilità e durata. I sistemi Well-Architected utilizzano più soluzioni di storage e attivano funzionalità diverse per migliorare le prestazioni e utilizzare le risorse in modo efficiente.

Reti e distribuzione di contenuti

La soluzione di rete ottimale per un carico di lavoro varia in base a latenza, requisiti di throughput, jitter e larghezza di banda. I vincoli fisici, ad esempio le risorse utente o on-premises, determinano le opzioni di posizione. Questi vincoli possono essere compensati con le posizioni edge o la collocazione delle risorse.

On AWS, la rete è virtualizzata ed è disponibile in diversi tipi e configurazioni. In questo modo è più facile soddisfare le esigenze di rete. AWS offre funzionalità di prodotto (ad esempio Enhanced Networking, istanze ottimizzate per la EC2 rete Amazon, accelerazione del trasferimento Amazon S3 e CloudFront Amazon dinamico) per ottimizzare il traffico di rete. AWS offre anche funzionalità di rete (ad esempio, routing di latenza Amazon Route 53 AWS Direct Connect, VPC endpoint Amazon e AWS Global Accelerator) per ridurre la distanza di rete o il jitter.

Le seguenti domande si concentrano su queste considerazioni relative all'efficienza delle prestazioni.

PERF4: Come selezionate e configurate le risorse di rete nel vostro carico di lavoro?

Questa domanda comprende linee guida e best practice per progettare, configurare e gestire soluzioni di rete e distribuzione di contenuti nel cloud in maniera efficiente.

Processo e cultura

Durante la fase di progettazione dei carichi di lavoro, esistono principi e pratiche che è possibile adottare per gestire al meglio carichi di lavoro cloud efficienti e ad alte prestazioni. Per adottare una

cultura che promuova l'efficienza delle prestazioni dei carichi di lavoro cloud, prendi in considerazione questi principi e pratiche fondamentali.

Per sviluppare questa cultura, considera questi principi chiave:

- Infrastruttura come codice: definisci l'infrastruttura come codice utilizzando approcci come i AWS CloudFormation modelli. L'uso dei modelli ti consente di collocare la tua infrastruttura nel controllo sorgente, insieme al codice e alle configurazioni dell'applicazione. Ciò ti premette di applicare le stesse procedure di sviluppo software all'infrastruttura, in modo da accelerare l'iterazione.
- Pipeline di implementazione: usa una pipeline di integrazione continua/implementazione continua (CI/CD) (ad esempio, repository del codice sorgente, sistemi di sviluppo, distribuzione e automazione dei test) per distribuire la tua infrastruttura. Ciò ti consente di effettuare l'implementazione in modo ripetibile, coerente ed economicamente vantaggioso nel corso dell'iterazione.
- Metriche ben definite: configura e monitora le metriche per acquisire gli indicatori chiave di prestazione (). KPIs Ti consigliamo di adottare parametri tecnici e aziendali. Per i siti Web o le app mobili, le metriche chiave sono l'acquisizione o il rendering. time-to-first-byte Gli altri parametri generalmente validi includono il numero di thread, il tasso di rimozione di oggetti inutili (garbage collection) e gli stati di attesa. I parametri aziendali, come il costo cumulativo aggregato per richiesta, possono indicarti due modi per ridurre i costi. Valuta attentamente il modo in cui prevedi di interpretare i parametri. Ad esempio, potresti scegliere il 99° percentile o quello massimo anziché il valore medio.
- Automatizza i test delle prestazioni: nell'ambito del processo di implementazione, avvia automaticamente i test delle prestazioni dopo che quelli dall'esecuzione più rapida hanno dato esito positivo. L'automazione deve creare un nuovo ambiente, configurare le condizioni iniziali come i dati del test ed eseguire una serie di benchmark e test di carico. I risultati dei test devono essere confrontati con la build, in modo da monitorare le variazioni delle prestazioni nel corso del tempo. Per i test di lunga durata, puoi inserirli nella pipeline in maniera asincrona rispetto al resto della build. In alternativa, puoi eseguire test delle prestazioni durante la notte utilizzando Amazon EC2 Spot Instances.
- Generazione del carico: crea una serie di script di test che replichino i percorsi utente sintetici o pre-registrati. Tali script devono essere idempotenti e non devono essere associati in coppie. Inoltre, potrebbe essere necessario includere script preliminari per garantire risultati validi. Testa gli script il più possibile, per assicurarti che replichino le abitudini di utilizzo in produzione. È possibile utilizzare soluzioni software o software-as-a-service (SaaS) per generare il carico. Valuta se

l'utilizzo delle soluzioni Marketplace AWS e le istanze spot possono essere modi convenienti per generare il carico.

- Visibilità delle prestazioni: i parametri principali devono essere visibili dal team, in particolar modo quelli relativi a ciascuna versione della build. Ciò ti consente di rilevare tendenze positive o negative rilevanti nel corso del tempo. Dovresti anche visualizzare i parametri sul numero di errori o eccezioni per assicurarti di testare un sistema funzionante.
- Visualizzazione: sfrutta le tecniche di visualizzazione che indicano in modo chiaro i punti in cui si verificano problemi di prestazioni, hot spot, stati di attesa o utilizzo ridotto. Sovrapponi i parametri delle prestazioni ai diagrammi architetturali: i grafici delle chiamate o il codice possono aiutarti a individuare più rapidamente i problemi.
- Revisione regolare dei processi: le prestazioni scarse delle architetture sono in genere il risultato di un processo di revisione delle prestazioni inesistente o incompleto. Se la tua architettura offre prestazioni insufficienti, l'implementazione di un processo di revisione delle prestazioni ti consente di favorire il miglioramento delle iterazioni.
- Ottimizzazione continua: adotta una cultura per ottimizzare continuamente l'efficienza delle prestazioni del tuo carico di lavoro cloud.

Le seguenti domande si concentrano su queste considerazioni relative all'efficienza delle prestazioni.

PERF5: Quale processo utilizzate per supportare una maggiore efficienza delle prestazioni per il vostro carico di lavoro?

Durante la fase di progettazione dei carichi di lavoro, esistono principi e pratiche che è possibile adottare per gestire al meglio carichi di lavoro cloud efficienti e ad alte prestazioni. Per adottare una cultura che promuova l'efficienza delle prestazioni dei carichi di lavoro cloud, prendi in considerazione questi principi e pratiche fondamentali.

Risorse

Consulta le seguenti risorse per ulteriori informazioni sulle best practice relative all'efficienza delle prestazioni.

Documentazione

- Ottimizzazione delle prestazioni di Amazon S3

- Prestazioni Amazon EBS Volume

Whitepaper

- Pilastro dell'efficienza delle prestazioni

Video

- AWS re:Invent 2019: EC2 fondamenti di Amazon (-R2) CMP211
- AWS re:Invent 2019: Sessione di leadership: Lo stato di archiviazione dell'unione (01-L) STG2
- AWS re:Invent 2019: Sessione di leadership: database creati appositamente (09-L) AWS DAT2
- AWS re:Invent 2019: Connettività e architetture di rete ibride (-R1) AWSAWS NET317
- AWS re:Invent 2019: Potenziamento di EC2 Amazon di nuova generazione: approfondimenti sul sistema Nitro (03-R2) CMP3
- AWS re:Invent 2019: scalabilità fino ai primi 10 milioni di utenti (-R) ARC211

Ottimizzazione dei costi

Il pilastro dell'ottimizzazione dei costi include la possibilità di eseguire sistemi per offrire valore aggiunto al prezzo più basso.

Il pilastro dell'ottimizzazione dei costi offre una panoramica dei principi di progettazione, delle best practice e delle domande. Le linee guida con le prescrizioni sull'implementazione sono disponibili nel whitepaper sul pilastro dell'ottimizzazione dei costi.

Argomenti

- Principi di progettazione
- Definizione
- Best practice
- Risorse

Principi di progettazione

I principi di progettazione per l'ottimizzazione dei costi nel cloud sono cinque:

- Implementa la gestione finanziaria del cloud: per migliorare i risultati finanziari e accelerare la realizzazione del valore aziendale nel cloud, investi nella gestione finanziaria e nell'ottimizzazione dei costi sul cloud. L'organizzazione deve dedicare tempo e risorse per creare capacità in questo nuovo dominio di gestione della tecnologia e dell'utilizzo. Come per le funzionalità di sicurezza o eccellenza operativa, la capacità si crea tramite lo sviluppo di competenze, programmi, risorse e processi destinati a far diventare un'organizzazione efficiente in termini di costi.
- Adotta un modello a consumo: paga solo le risorse di calcolo che richiedi e incrementa o riduci l'utilizzo a seconda delle reali necessità aziendali anziché sulla base di complesse previsioni. Ad esempio, gli ambienti di test e di sviluppo sono in genere usati solo per otto ore al giorno durante la settimana lavorativa. Puoi interrompere queste risorse quando non le utilizzi, risparmiando potenzialmente il 75% dei costi (40 ore anziché 168).
- Misura l'efficienza complessiva: misura i risultati aziendali del carico di lavoro e i costi associati alla sua fornitura. Usa questi dati per determinare i ricavi che puoi ottenere dal miglioramento dei risultati e dalla riduzione dei costi.
- Smettila di spendere soldi per il sollevamento indifferenziato di carichi pesanti: AWS si occupa del lavoro gravoso delle operazioni del data center come scaffalature, impilamento e alimentazione dei server. Inoltre, elimina l'onere operativo della gestione di sistemi operativi e applicazioni con servizi gestiti. In questo modo, potrai dedicarti ai clienti e ai progetti aziendali anziché dell'infrastruttura IT.
- Analizza e attribuisci le spese: il cloud semplifica l'individuazione precisa di utilizzo e costo dei sistemi, favorendo l'attribuzione trasparente dei costi IT ai singoli proprietari dei carichi di lavoro. Questo aiuta a misurare il ritorno sull'investimento (ROI) e offre ai proprietari dei carichi di lavoro l'opportunità di ottimizzare le proprie risorse e ridurre i costi.

Definizione

Esistono cinque aree di best practice per l'ottimizzazione dei costi nel cloud:

- Implementazione della gestione finanziaria del cloud
- Comprensione delle spese e dell'utilizzo
- Risorse convenienti in termini di costo
- Gestione delle risorse di domanda e offerta
- Ottimizzazione nel tempo

Come per gli altri pilastri del Well-Architected Framework, ci sono dei compromessi da considerare, ad esempio, se ottimizzare i costi o i costi. speed-to-market In alcuni casi, è più efficiente ottimizzare

la velocità, per velocizzare il lancio sul mercato, fornire nuove funzionalità o rispettare una scadenza, anziché investire nell'ottimizzazione dei costi iniziali. Talvolta le decisioni di progettazione sono guidate dalla rapidità invece che dai dati, ed esiste sempre la tentazione di sovrascrivere piuttosto che dedicare tempo all'esecuzione di benchmark per la implementazione più conveniente. Questo potrebbe portare a implementazione con provisioning eccessivo e sottoutilizzate. Tuttavia, si tratta di una scelta ragionevole quando devi eseguire il "lift and shift" delle risorse dal tuo ambiente on-premises al cloud e procedere con l'ottimizzazione di conseguenza. Investire in anticipo la giusta quantità di energia in una strategia di ottimizzazione dei costi permette di realizzare i vantaggi economici del cloud in modo più rapido, ottenendo il rispetto costante delle best practice ed evitando un provisioning eccessivo. Le sezioni seguenti forniscono tecniche e best practice per l'implementazione iniziale e continua della gestione finanziaria del cloud e l'ottimizzazione dei costi dei carichi di lavoro.

Best practice

Argomenti

Implementazione della gestione finanziaria del cloud

Con l'adozione del cloud, i team tecnologici innovano più rapidamente grazie a cicli di approvazione, approvvigionamento e implementazione dell'infrastruttura più brevi. Per ottenere valore aggiunto e migliorare gli affari è necessario un nuovo approccio alla gestione finanziaria nel cloud. Questo approccio è la gestione finanziaria del cloud e crea capacità in tutta l'organizzazione implementando competenze, programmi, risorse e processi a livello organizzativo.

Molte organizzazioni sono composte da tante unità con priorità diverse. La capacità di allineare un'organizzazione a un insieme concordato di obiettivi finanziari e di fornire all'organizzazione i meccanismi per raggiungerli permette di creare un'organizzazione più efficiente. Un'organizzazione capace innova e crea più rapidamente, è più agile e si adatta a qualsiasi fattore interno o esterno.

AWS Puoi utilizzare Cost Explorer e, facoltativamente, Amazon Athena e QuickSight Amazon con il Cost and Usage Report CUR (), per fornire informazioni su costi e utilizzo in tutta l'organizzazione.

AWS Budgets fornisce notifiche proattive relative a costi e utilizzo. I AWS blog forniscono informazioni su nuovi servizi e funzionalità per verificare che tu sia sempre aggiornato sulle nuove versioni dei servizi.

La seguente domanda si concentra su queste considerazioni relative all'ottimizzazione dei costi. Per l'elenco completo delle domande e delle best practice relative all'ottimizzazione dei costi, consulta l'Appendice.

COST1: Come si implementa la gestione finanziaria del cloud?

L'implementazione del Cloud Financial Management aiuta le organizzazioni a realizzare valore aziendale e successo finanziario ottimizzando i costi, l'utilizzo e la scalabilità AWS.

Quando crei una funzione di ottimizzazione dei costi, utilizza membri e integra il team con esperti nell'CFMottimizzazione dei costi. I membri già presenti nel team conoscono il funzionamento dell'organizzazione e sono in grado di implementare rapidamente i miglioramenti. Valuta anche la possibilità di includere persone con competenze aggiuntive o specialistiche, ad esempio di analisi e gestione dei progetti.

Quando implementi la consapevolezza dei costi nella tua organizzazione, prova a migliorare o sviluppare i programmi e i processi esistenti. È molto più veloce sviluppare i processi e programmi esistenti, piuttosto che crearne di nuovi. In questo modo puoi ottenere risultati molto più rapidamente.

Comprensione delle spese e dell'utilizzo

La maggiore flessibilità e agilità consentite dal cloud incoraggiano l'innovazione, lo sviluppo e l'implementazione rapidi. Riduce i processi manuali e il tempo associati al provisioning dell'infrastruttura on-premises, tra cui l'identificazione delle specifiche hardware, la negoziazione delle quotazioni dei prezzi, la gestione degli ordini di acquisto, la pianificazione delle spedizioni e la distribuzione delle risorse. Tuttavia, la facilità d'uso e la capacità on demand virtualmente illimitata richiedono un nuovo tipo di mentalità in merito alle spese.

Molte aziende sono caratterizzate da più sistemi gestiti da vari team. La capacità di attribuire i costi delle risorse ai singoli proprietari dell'organizzazione o del prodotto incoraggia un comportamento di utilizzo efficiente e contribuisce a ridurre gli sprechi. L'attribuzione precisa dei costi consente di capire quali prodotti sono effettivamente redditizi e permette anche di prendere decisioni più consapevoli in merito alle destinazioni del budget.

In AWS, crei una struttura contabile con AWS Organizations or AWS Control Tower, che fornisce la separazione e aiuta nell'allocazione dei costi e dell'utilizzo. Puoi anche utilizzare l'applicazione di tag alle risorse per associare informazioni aziendali e organizzative a utilizzo e costi. AWS Cost Explorer Utilizzalo per avere visibilità su costi e utilizzo oppure crea dashboard e analisi personalizzate con Amazon Athena e Amazon. QuickSight Il controllo dei costi e dell'utilizzo avviene tramite notifiche tramite AWS Budgets e controlli tramite AWS Identity and Access Management (IAM) e Service Quotas.

Le seguenti domande si concentrano su queste considerazioni relative all'ottimizzazione dei costi.

COST2: Come regolate l'utilizzo?

Stabilisci policy e meccanismi per convalidare che i costi sostenuti mentre raggiungi gli obiettivi siano adeguati. Utilizzando un checks-and-balances approccio, è possibile innovare senza spendere troppo.

COST3: Come monitorate l'utilizzo e i costi?

Stabilisci policy e procedure per monitorare e allocare i costi in modo appropriato. Ciò ti permette di misurare e migliorare l'efficienza in termini di costi del carico di lavoro.

COST4: Come si disattivano le risorse?

Implementa il controllo delle modifiche e la gestione delle risorse dall'inizio del progetto fino a. end-of-life In questo modo sarà più semplice disattivare le risorse inutilizzate per ridurre gli sprechi.

Puoi usare i tag di allocazione dei costi per categorizzare e monitorare il tuo utilizzo di AWS e i costi. Quando applichi tag alle tue AWS risorse (come EC2 istanze o bucket S3), AWS genera un rapporto sui costi e sull'utilizzo con l'utilizzo e i tag. Puoi applicare tag che rappresentano le categorie di un'organizzazione (come i centri di costo, i nomi dei carichi di lavoro o i proprietari) per organizzare i tuoi costi tra i vari servizi.

Verifica di utilizzare il giusto livello di dettaglio e granularità quando crei report e monitori costi e utilizzo. Per informazioni e tendenze generali, utilizza i dati giornalieri di AWS Cost Explorer. Per un'analisi e un'ispezione più approfondite AWS Cost Explorer, utilizza la granularità oraria in o

Amazon Athena e Amazon QuickSight con il rapporto sui costi e l'utilizzo (CUR) con una granularità oraria.

Associando le risorse taggate al monitoraggio del ciclo di vita dell'entità (dipendenti, progetti), puoi individuare le risorse accantonate o i progetti che non generano più valore per l'organizzazione e devono quindi essere dismessi. Puoi impostare avvisi di fatturazione per ricevere notifiche relative a spese eccessive previste.

Risorse convenienti in termini di costo

Utilizzare risorse e istanze adeguate al tuo carico di lavoro è fondamentale per ridurre i costi. Ad esempio, un processo di creazione di report potrebbe impiegare cinque ore su un server più piccolo, ma un'ora su un server più grande che costa il doppio. Entrambi i server ti offrono lo stesso risultato, ma quello più piccolo comporta un costo più elevato nel tempo.

Un carico di lavoro ben progettato usa le risorse più convenienti, il che può avere un impatto economico positivo e notevole. Hai anche la possibilità di usare i servizi gestiti per ridurre i costi. Ad esempio, invece di mantenere dei server per recapitare le e-mail, puoi usare un servizio che ti invia gli addebiti in base ai messaggi inviati.

AWS offre una varietà di opzioni di prezzo flessibili ed economiche per acquistare istanze da Amazon EC2 e altri servizi in modo più adatto alle tue esigenze. Le istanze on demand ti permettono di pagare la capacità di calcolo a ore e non richiedono impegni minimi. Savings Plans e istanze riservare garantiscono risparmi fino al 75% sui prezzi delle istanze on demand. Con le istanze Spot, puoi sfruttare la capacità EC2 Amazon inutilizzata e offrire risparmi fino al 90% sui prezzi On-Demand. Le istanze Spot sono adatte laddove il sistema può tollerare l'utilizzo di una flotta di server in cui i singoli server possono entrare e uscire in modo dinamico, come server Web stateless, elaborazione in batch o quando si utilizzano big data. HPC

Una selezione appropriata dei servizi può anche ridurre l'utilizzo e i costi, ad esempio riducendo CloudFront al minimo il trasferimento dei dati o diminuendo i costi, ad esempio utilizzando Amazon Aurora su RDS Amazon per rimuovere i costosi costi di licenza dei database.

Le seguenti domande si concentrano su queste considerazioni relative all'ottimizzazione dei costi.

COST5: Come si valutano i costi quando si selezionano i servizi?

Amazon EC2EBS, Amazon e Amazon S3 sono servizi integrati AWS . I servizi gestiti, come Amazon RDS e Amazon DynamoDB, sono servizi di livello superiore o a livello di applicazione.

COST5: Come si valutano i costi quando si selezionano i servizi?

AWS Selezionando i blocchi predefiniti e i servizi gestiti appropriati, è possibile ottimizzare questo carico di lavoro per i costi. Ad esempio, utilizzando i servizi gestiti, puoi ridurre o eliminare gran parte dei costi generali amministrativi e operativi, liberandotene per lavorare su applicazioni e attività correlate al tuo business.

COST6: Come si raggiungono gli obiettivi di costo selezionando il tipo, la dimensione e il numero di risorse?

Assicurati di scegliere la dimensione e il numero delle risorse appropriati per l'attività in questione. Selezionando il tipo, le dimensioni e il numero più convenienti, riduci al minimo gli sprechi.

COST7: Come si utilizzano i modelli di prezzo per ridurre i costi?

Utilizza il modello di prezzo più appropriato per le tue risorse per ridurre al minimo le spese.

COST8: Come pianificate i costi di trasferimento dei dati?

Assicurati di pianificare e monitorare i costi di trasferimento dei dati in modo da poter prendere decisioni sull'architettura per ridurre al minimo i costi. Una modifica piccola ma efficace dell'arch itettura può ridurre drasticamente i costi operativi nel tempo.

Tenendo conto dei costi durante la selezione del servizio e utilizzando strumenti come Cost Explorer e AWS Trusted Advisor controllando regolarmente AWS l'utilizzo, è possibile monitorare attivamente l'utilizzo e adattare di conseguenza le implementazioni.

Gestione delle risorse di domanda e offerta

Quando passi al cloud, paghi solo ciò che ti occorre. Puoi fornire risorse in base alla domanda del carico di lavoro nel momento in cui sono necessarie, riducendo così la necessità di un provisioning eccessivo, costoso e dispendioso. Puoi anche gestire la domanda utilizzando tecniche come limitazione (della larghezza di banda della rete), buffering o queuing per allentare la domanda e soddisfarla con meno risorse. In questo modo diminuirai i costi o li posticiperai con un servizio batch.

In AWS, è possibile fornire automaticamente le risorse in base alla domanda del carico di lavoro. Auto Scaling con strategie basate su domanda o tempo ti permette di aggiungere e rimuovere le risorse in base alle esigenze. Se riesci a prevedere le variazioni nella domanda, puoi risparmiare di più e convalidare che le risorse corrispondano alle esigenze del tuo carico di lavoro. Puoi utilizzare Amazon API Gateway per implementare la limitazione o Amazon SQS per implementare una coda nel tuo carico di lavoro. Entrambi permettono di modificare la richiesta nei componenti del carico di lavoro.

La seguente domanda si concentra su queste considerazioni relative all'ottimizzazione dei costi.

COST9: Come gestisci la domanda e l'offerta di risorse?

Per avere un carico di lavoro con costo e prestazioni bilanciate, verifica che venga utilizzato tutto ciò per cui paghi ed evita le istanze molto sottoutilizzate. Una metrica di utilizzo distorta in entrambe le direzioni ha un impatto negativo sull'organizzazione, sia sui costi operativi (riduzion e delle prestazioni dovuta all'eccessivo utilizzo) sia sugli sprechi di spesa (dovuti all'eccessivo approvvigionamento). AWS

Quando progetti di modificare le risorse di domanda e offerta, pensa attentamente ai modelli di utilizzo, al tempo necessario per effettuare il provisioning delle nuove risorse e alla prevedibilità del modello di domanda. Quando gestisci la domanda, verifica di disporre di una coda o di un buffer di dimensioni corrette e di rispondere alla domanda del carico di lavoro nel periodo di tempo richiesto.

Ottimizzazione nel tempo

Non appena vengono AWS rilasciati nuovi servizi e funzionalità, è consigliabile rivedere le decisioni architettoniche esistenti per verificare che continuino a essere le più convenienti. Man mano che le tue esigenze cambiano, disattiva tempestivamente risorse, interi servizi e sistemi non appena smettono di essere necessari.

L'implementazione di nuove funzionalità o tipi di risorse può ottimizzare il carico di lavoro in modo incrementale e con uno sforzo minimo. In questo modo puoi migliorare continuamente l'efficienza nel tempo e utilizzare le tecnologie più aggiornate per ridurre i costi operativi. Puoi anche sostituire o aggiungere nuovi componenti al carico di lavoro con nuovi servizi. In questo modo puoi aumentare in modo significativo l'efficienza, perciò è essenziale rivedere regolarmente il carico di lavoro e implementare nuovi servizi e caratteristiche.

Le seguenti domande si concentrano su queste considerazioni relative all'ottimizzazione dei costi.

COST10: Come valutate i nuovi servizi?

Non appena vengono AWS rilasciati nuovi servizi e funzionalità, è consigliabile rivedere le decisioni architettoniche esistenti per verificare che continuino a essere le più convenienti.

Quando esamini regolarmente le tue implementazioni, valuta in che modo i servizi più recenti possono aiutarti a risparmiare. Ad esempio, Amazon Aurora su Amazon RDS può ridurre i costi per i database relazionali. L'utilizzo di serverless come Lambda consente di eliminare la necessità di utilizzare e gestire le istanze per eseguire il codice.

COST11: Come valuti il costo dell'impegno?

Valuta il costo delle operazioni nel cloud, esamina le operazioni cloud che richiedono molto tempo e automatizzale per ridurre gli sforzi umani e i costi adottando AWS servizi correlati, prodotti di terze parti o strumenti personalizzati.

Risorse

Consulta le seguenti risorse per ulteriori informazioni sulle nostre best practice per l'ottimizzazione dei costi.

Documentazione

- Documentazione AWS

Whitepaper

- Cost Optimization Pillar

Sostenibilità

Alla base del pilastro della sostenibilità c'è l'attenzione all'impatto ambientale, soprattutto in termini di uso ed efficienza delle fonti energetiche, leve importanti che gli architetti usano per definire interventi diretti mirati a ridurre lo sfruttamento delle risorse. Puoi trovare linee guida prescrittive sull'implementazione nel whitepaper sul pilastro della sostenibilità.

Argomenti

Principi di progettazione

Esistono sei principi di progettazione per la sostenibilità nel cloud:

- Analizza il tuo impatto: misura l'impatto del tuo carico di lavoro cloud e definiscine l'impatto futuro. Nella tua analisi includi ogni fonte di impatto: quelle derivanti dall'uso dei prodotti da parte dei tuoi clienti e quelle derivanti dalla rimozione e dal ritiro finali dal mercato. Confronta l'output di produzione e l'impatto totale dei tuoi carichi di lavoro cloud, partendo dall'analisi di risorse ed emissioni richieste per unità di lavoro. Utilizzate questi dati per stabilire gli indicatori chiave di prestazione (KPIs), valutare i modi per migliorare la produttività riducendo l'impatto e stimare l'impatto delle modifiche proposte nel tempo.
- Stabilisci obiettivi di sostenibilità: per ciascun carico di lavoro cloud, stabilisci obiettivi di sostenibilità a lungo termine, come, ad esempio, ridurre le risorse di calcolo e di archiviazione richieste per ciascuna transazione. Modella il ritorno sugli investimenti finalizzati alle migliorie in materia di sostenibilità per i carichi di lavoro esistenti e offri ai proprietari le risorse di cui hanno bisogno per investire negli obiettivi di sostenibilità. Pianifica lo sviluppo e progetta i tuoi carichi di lavoro in modo che la crescita comporti un impatto meno intenso se misurato rispetto a un'unità appropriata, come l'utente o la transazione. Gli obiettivi ti aiutano ad avvalorare un progetto più ampio di sostenibilità che coinvolge la tua azienda o la tua organizzazione, a identificare le regressioni e a dare la priorità a quelle aree che offrono un maggiore potenziale di miglioramento.
- Massimizza l'utilizzo: dimensiona correttamente i carichi di lavoro e implementa un progetto per verificare un utilizzo elevato e ottimizzare l'efficienza energetica dell'hardware sottostante. Due host in esecuzione con una percentuale di utilizzo pari al 30% sono meno efficienti di un host in esecuzione al 60%, se consideriamo il consumo di base per host. Allo stesso tempo, elimina o riduci le risorse, le elaborazioni e le archiviazioni inattive per ridurre l'energia totale richiesta per alimentare il tuo carico di lavoro.
- Anticipa e adotta nuove offerte hardware e software più efficienti: supporta i miglioramenti a monte apportati dai tuoi partner e fornitori così da ridurre l'impatto dei tuoi carichi di lavoro sul cloud.

Monitora costantemente il mercato e valuta nuove offerte hardware e software più efficienti. Adotta la flessibilità nei tuoi progetti per consentire una rapida adozione di tecnologie nuove ed efficienti.

- Affidati a servizi gestiti: la condivisione dei servizi con un'ampia base clienti consente di ottimizzare l'uso delle risorse e ridurre al tempo stesso l'infrastruttura necessaria per supportare i carichi di lavoro nel cloud. Ad esempio, i clienti possono condividere l'impatto dei componenti comuni dei data center come l'alimentazione e la rete migrando i carichi di lavoro verso Cloud AWS e adottando servizi gestiti, come AWS Fargate per i container serverless, che AWS opera su larga scala ed è responsabile del loro funzionamento efficiente. Utilizza servizi gestiti che possono contribuire a ridurre al minimo l'impatto, come lo spostamento automatico dei dati a cui si accede raramente in cold storage con configurazioni del ciclo di vita di Amazon S3 o Amazon EC2 Auto Scaling per regolare la capacità in base alla domanda.
- Riduci l'impatto a valle dei carichi di lavoro nel cloud: riduci la quantità di energia o di risorse impiegate nell'utilizzo dei tuoi servizi. Riduci la necessità di eseguire aggiornamenti dei tuoi dispositivi per usare i tuoi servizi. Esegui test usando device farm per analizzare l'impatto atteso e conduci altri test con i clienti per capire l'impatto reale derivante dall'uso dei tuoi servizi.

Definizione

Esistono sei aree di best practice per la sostenibilità nel cloud:

- Selezione della regione
- Allineamento alla domanda
- Software e architettura
- Dati
- Hardware e servizi
- Processo e cultura

Sostenibilità nel cloud significa impegnarsi continuamente per ridurre principalmente il consumo di energia e garantire una maggiore efficienza di tutti i componenti di un carico di lavoro, ottenendo il massimo vantaggio dalle risorse allocate e riducendo al minimo le risorse richieste. Tale impegno va dalla selezione iniziale di un linguaggio di programmazione efficace, dall'adozione di algoritmi moderni e dall'uso di tecniche di archiviazione di dati efficienti alla distribuzione in infrastrutture di calcolo valide e correttamente dimensionate e alla riduzione dei requisiti per l'hardware degli utenti finali a potenza elevata.

Best practice

Argomenti

Selezione della regione

La scelta della regione per il carico di lavoro influisce in modo significativo su prestazioniKPIs, costi e impronta di carbonio. Per migliorarliKPIs, dovresti scegliere le regioni per i tuoi carichi di lavoro in base ai requisiti aziendali e agli obiettivi di sostenibilità.

La seguente domanda si concentra su queste considerazioni relative alla sostenibilità. Per l'elenco completo delle domande e delle best practice relative all'affidabilità, consulta l'Appendice.

SUS1: Come selezionate le regioni per il vostro carico di lavoro?

La scelta della regione per il carico di lavoro influisce in modo significativo sul carico di lavoroKPI s, in termini di prestazioni, costi e impronta di carbonio. Per migliorarliKPIs, dovresti scegliere le regioni per i tuoi carichi di lavoro in base ai requisiti aziendali e agli obiettivi di sostenibilità.

Allineamento alla domanda

Il modo in cui gli utenti e le applicazioni utilizzano i tuoi carichi di lavoro e altre risorse può aiutarti a identificare i miglioramenti da implementare per raggiungere gli obiettivi di sostenibilità. Puoi scalare l'infrastruttura in modo che sia costantemente adatta alla domanda e verifica di usare solo le risorse minime necessarie per supportare gli utenti. Allinea i livelli di servizio alle esigenze dei clienti. Colloca le risorse in modo da limitare la rete necessaria per il loro consumo da parte di utenti e applicazioni. Rimuovi gli asset inutilizzati. Offri ai membri del team dispositivi in grado di soddisfarne le esigenze con un impatto minimo in termini di sostenibilità.

La seguente domanda si concentra su queste considerazioni relative alla sostenibilità:

SUS2: Come allineate le risorse cloud alla vostra richiesta?

Il modo in cui gli utenti e le applicazioni utilizzano i tuoi carichi di lavoro e altre risorse può aiutarti a identificare i miglioramenti da implementare per raggiungere gli obiettivi di sostenibilità. Puoi scalare l'infrastruttura in modo che sia costantemente adatta alla domanda e verifica di usare solo le risorse minime necessarie per supportare gli utenti. Allinea i livelli di servizio alle esigenze dei clienti. Colloca le risorse in modo da limitare la rete necessaria per il loro consumo da parte di utenti e applicazioni. Rimuovi gli asset inutilizzati. Offri ai membri del team dispositivi in grado di soddisfarne le esigenze con un impatto minimo in termini di sostenibilità.

Dimensiona l'infrastruttura in base al carico degli utenti: identifica i periodi di utilizzo assente o ridotto e scala le risorse per ridurre capacità in eccesso e migliorare l'efficienza.

Allineamento SLAs agli obiettivi di sostenibilità: definisci e aggiorna gli accordi sui livelli di servizio (SLAs), come la disponibilità o i periodi di conservazione dei dati, per ridurre al minimo il numero di risorse necessarie per supportare il carico di lavoro continuando a soddisfare i requisiti aziendali.

Riduci la creazione e la manutenzione di asset inutilizzati: analizza le risorse delle applicazioni (come report precompilati, set di dati e immagini statiche) e i modelli di accesso alle risorse per identificare ridondanze, sottoutilizzi e obiettivi potenziali di disattivazione. Consolida le risorse generate con contenuti ridondanti (come, ad esempio, report mensili con set di dati e output comuni o in sovrapposizione) per ridurre le risorse utilizzate per la duplicazione degli output. Disattiva le risorse non utilizzate (come, ad esempio, immagini di prodotto non più in vendita) per rilasciare le risorse usate e ridurre il numero di risorse sfruttate per supportare il carico di lavoro.

Ottimizza il posizionamento geografico dei carichi di lavoro in base alle posizioni degli utenti: analizza i modelli di accesso alla rete per capire da quali aree geografiche si connettono i tuoi clienti. Seleziona le regioni e i servizi per ridurre la distanza che il traffico di rete deve percorrere e diminuire così le risorse totali di rete richieste per supportare il tuo carico di lavoro.

Ottimizza le risorse dei membri del team in base alle attività eseguite: ottimizza le risorse fornite ai membri del team per ridurre al minimo l'impatto sulla sostenibilità e supportare al tempo stesso le loro esigenze. Esegui ad esempio operazioni complesse, come rendering e compilazione, su desktop cloud condivisi altamente usati invece che su sistemi per utenti singoli, sottoutilizzati e con un alto dispendio energetico.

Software e architettura

Implementa modelli per eseguire lo smoothing del carico e garantire un utilizzo elevato e coerente delle risorse implementate per ridurre al minimo il loro consumo. In seguito alle modifiche nei comportamenti degli utenti nel tempo, alcuni componenti potrebbero diventare inattivi per mancanza di utilizzo. Rivedi modelli e architetture per consolidare i componenti sottoutilizzati e aumentare l'uso complessivo. Ritira i componenti non più necessari. Analizza le prestazioni dei componenti dei tuoi carichi di lavoro e ottimizza quelli che usano la maggior quantità di risorse. Identifica i dispositivi che i clienti utilizzano per accedere ai servizi e implementa modelli in grado di ridurre al minimo la necessità di aggiornamenti dei dispositivi.

La seguente domanda si concentra su queste considerazioni relative alla sostenibilità:

SUS3: Come sfruttate i modelli di software e architettura per supportare i vostri obiettivi di sostenibilità?

Implementa modelli per eseguire lo smoothing del carico e garantire un utilizzo elevato e coerente delle risorse implementate per ridurre al minimo il loro consumo. In seguito alle modifiche nei comportamenti degli utenti nel tempo, alcuni componenti potrebbero diventare inattivi per mancanza di utilizzo. Rivedi modelli e architetture per consolidare i componenti sottoutilizzati e aumentare l'uso complessivo. Ritira i componenti non più necessari. Analizza le prestazioni dei componenti dei tuoi carichi di lavoro e ottimizza quelli che usano la maggior quantità di risorse. Identifica i dispositivi che i clienti utilizzano per accedere ai servizi e implementa modelli in grado di ridurre al minimo la necessità di aggiornamenti dei dispositivi.

Ottimizza software e architetture per processi asincroni e pianificati: utilizza progettazioni e architetture software efficienti per ridurre al minimo le risorse medie richieste per unità di lavoro. Implementa meccanismi che generano un utilizzo uniforme dei componenti per ridurre le risorse inattive tra le attività e diminuire l'impatto di picchi di carico.

Rimuovi o rifattorizza i componenti dei carichi di lavoro con un utilizzo ridotto o assente: monitora l'attività dei carichi di lavoro per individuare i cambiamenti che si verificano nel tempo nell'utilizzo dei singoli componenti. Elimina i componenti non utilizzati e non più necessari e procedi a rifattorizzare quelli con scarso utilizzo per limitare lo spreco di risorse.

Ottimizza le aree di codice che consumano la maggior parte del tempo o delle risorse: monitora l'attività dei carichi di lavoro per individuare i componenti delle applicazioni che usano la maggior

parte delle risorse. Ottimizza il codice eseguito all'interno di questi componenti per ridurre l'utilizzo delle risorse e massimizzare al tempo stesso le prestazioni.

Ottimizza l'impatto su dispositivi e apparecchiature dei clienti: identifica i dispositivi e le attrezzature che i tuoi clienti usano per accedere ai tuoi servizi, il loro ciclo di vita atteso e l'impatto finanziario e di sostenibilità che deriva dalla loro sostituzione. Implementa modelli e architetture software per ridurre al minimo la necessità dei clienti di sostituire dispositivi e aggiornare attrezzature. Implementa ad esempio nuove caratteristiche usando un codice compatibile con versioni di hardware e sistemi operativi precedenti o gestisci la dimensione dei payload in modo che non superino la capacità di archiviazione del dispositivo target.

Usa i modelli e le architetture software che supportano al meglio l'accesso ai dati e i modelli di archiviazione: scopri come i dati vengono utilizzati all'interno del tuo carico di lavoro, consumati dagli utenti, trasferiti e archiviati. Seleziona tecnologie che ti consentono di ridurre l'elaborazione dei dati e i requisiti di archiviazione.

Gestione dei dati

La seguente domanda si concentra su queste considerazioni relative alla sostenibilità:

SUS4: Come sfruttate le politiche e i modelli di gestione dei dati per supportare i vostri obiettivi di sostenibilità?

Implementa procedure di gestione dei dati per ridurre l'archiviazione allocata richiesta per supportare il carico di lavoro e le risorse necessarie per l'uso correlato. Analizza i tuoi dati e usa tecnologie e configurazioni che supportano nel modo più efficace il valore aziendale dei dati e il modo in cui vengono utilizzati. Esegui il ciclo di vita dei dati su un'archiviazione più efficiente e meno performante al diminuire dei requisiti ed elimina i dati che non sono più necessari.

Implementa una policy di classificazione dei dati: classifica i dati per comprenderne il significato in favore dei risultati aziendali. Usa queste informazioni per stabilire quando trasferire i dati in un'archiviazione più efficiente dal punto di vista energetico o eliminarli in totale sicurezza.

Utilizza tecnologie che supportano l'accesso ai dati e i modelli di archiviazione: usa l'archiviazione in grado di supportare nel modo più efficiente il modo in cui viene effettuato l'accesso ai dati e come vengono archiviati per ridurre la quantità di risorse allocate e supportare al tempo stesso il tuo carico di lavoro. Ad esempio, i dispositivi a stato solido (SSDs) consumano più energia rispetto alle unità

magnetiche e devono essere utilizzati solo per casi di utilizzo attivo dei dati. Usa storage di classe di archiviazione ad alta efficienza energetica per i dati ad accesso infrequente.

Utilizza le policy del ciclo di vita per eliminare i dati non necessari: gestisci il ciclo di vita di tutti i tuoi dati e applica in automatico cronologie di eliminazione per ridurre i requisiti totali di archiviazione del tuo carico di lavoro.

Riduci il provisioning eccessivo nell'archiviazione a blocchi: per ridurre la quantità totale di archiviazione assegnata, crea un'archiviazione a blocchi con l'allocazione di dimensioni in base al carico di lavoro. Usa i volumi elastici per espandere l'archiviazione all'aumentare dei dati senza dover ridimensionare l'archiviazione collegata alle risorse di calcolo. Esamina regolarmente i volumi elastici e riduci i volumi con un provisioning eccessivo per adattarli alla dimensione corrente dei dati.

Elimina i dati ridondanti o non necessari: duplica i dati solo quando è necessario per ridurre la quantità totale di archiviazione utilizzata. Utilizza tecnologie di backup che deduplicano i dati a livello di file e blocco. Limita l'uso di configurazioni Redundant Array of Independent Drives (RAID), tranne laddove richiesto per soddisfare i requisiti. SLAs

Utilizza file system condivisi o archiviazione di oggetti per accedere a dati comuni: adotta l'archiviazione condivisa e singole fonti di verità per evitare la duplicazione dei dati e ridurre i requisiti di archiviazione complessiva del tuo carico di lavoro. Recupera i dati dall'archiviazione condivisa solo in base alle esigenze. Distacca volumi non utilizzati per rilasciare le risorse. Riduci al minimo gli spostamenti dei dati tra le reti: usa un'archiviazione condivisa e accedi ai dati da archivi regionali per contenere le risorse di rete totali necessarie per supportare i trasferimenti dei dati per il carico di lavoro.

Esegui il backup dei dati solo quando sono difficili da ricreare: per ridurre al minimo l'uso delle risorse di archiviazione, esegui il backup solo dei dati che abbiano un valore aziendale o siano considerati necessari per soddisfare requisiti di conformità. Esamina le policy di backup ed escludi l'archiviazione temporanea che non offre valore in uno scenario di ripristino.

Hardware e servizi

Cerca opportunità per ridurre l'impatto dei carichi di lavoro in termini di sostenibilità apportando modifiche alle tue pratiche di gestione hardware. Riduci al minimo la quantità di hardware necessaria per il provisioning e l'implementazione e scegli l'hardware e i servizi più efficienti per il singolo carico di lavoro.

La seguente domanda si concentra su queste considerazioni relative alla sostenibilità:

SUS5: Come selezionate e utilizzate l'hardware e i servizi cloud nella vostra architettura per supportare i vostri obiettivi di sostenibilità?

Cerca opportunità per ridurre l'impatto dei carichi di lavoro in termini di sostenibilità apportand o modifiche alle tue pratiche di gestione hardware. Riduci al minimo la quantità di hardware necessaria per il provisioning e l'implementazione e scegli l'hardware e i servizi più efficienti per il singolo carico di lavoro.

Utilizza la quantità minima di hardware per soddisfare le tue esigenze: le funzionalità del cloud consentono di apportare modifiche frequenti alle implementazioni dei carichi di lavoro. Aggiorna i componenti distribuiti man mano che le tue esigenze cambiano.

Usa tipi di istanze con il minimo impatto: monitora costantemente il rilascio di nuovi tipi di istanza e sfrutta le migliorie in tema di efficienza energetica, inclusi i tipi di istanza progettati per supportare carichi di lavoro specifici, come la formazione del machine learning, le inferenze e la transcodifica dei video.

Utilizza i servizi gestiti: i servizi gestiti trasferiscono la responsabilità del mantenimento di un utilizzo medio elevato e dell'ottimizzazione della sostenibilità dell'hardware distribuito su. AWS Utilizza i servizi gestiti per distribuire l'impatto della sostenibilità dei servizi su tutti i tenant relativi, riducendo così il singolo contributo.

Ottimizza l'uso diGPUs: Le unità di elaborazione grafica (GPUs) possono essere una fonte di elevato consumo energetico e molti GPU carichi di lavoro sono estremamente variabili, come il rendering, la transcodifica e la formazione e la modellazione tramite apprendimento automatico. Esegui GPUs le istanze solo per il tempo necessario e disattivale con l'automazione quando non è necessario per ridurre al minimo il consumo di risorse.

Processo e cultura

Cerca opportunità per ridurre l'impatto di sostenibilità apportando modifiche alle tue prassi di sviluppo, test e implementazione.

La seguente domanda si concentra su queste considerazioni relative alla sostenibilità:

SUS6: In che modo i vostri processi organizzativi supportano i vostri obiettivi di sostenibilità?

Cerca opportunità per ridurre l'impatto di sostenibilità apportando modifiche alle tue prassi di sviluppo, test e implementazione.

Adotta operazioni che consentono di integrare rapidamente i miglioramenti orientati alla sostenibilità: testa e convalida potenziali miglioramenti prima di distribuirli in produzione. Tieni in considerazione il costo dei test quando calcoli il potenziale vantaggio futuro di un miglioramento. Sviluppa operazioni di test a basso costo per agevolare la distribuzione di piccoli miglioramenti.

Mantieni aggiornato il tuo carico di lavoro: i sistemi Up-to-date operativi, le librerie e le applicazioni possono migliorare l'efficienza del carico di lavoro e favorire l'adozione di tecnologie più efficienti. Up-to-dateil software potrebbe anche includere funzionalità per misurare l'impatto sulla sostenibilità del carico di lavoro in modo più accurato, poiché i fornitori forniscono funzionalità per raggiungere i propri obiettivi di sostenibilità.

Incrementa l'utilizzo degli ambienti di sviluppo: utilizza l'automazione e il modello Infrastructure as code per rendere operativi gli ambienti di preproduzione quando necessario e dismetterli quando non vengono utilizzati. Un modello comune consiste nel pianificare periodi di disponibilità che coincidano con l'orario di lavoro dei membri del team incaricati dello sviluppo. L'ibernazione è uno strumento utile per preservare lo stato e portare rapidamente le istanze online solo quando necessario. Utilizza tipi di istanze con capacità di espansione, istanze spot, servizi di database elastici, container e altre tecnologie per allineare la capacità di sviluppo e test all'uso.

Utilizza device farm gestite per i test: le device farm gestiste distribuiscono l'impatto di sostenibilità della produzione di hardware e dell'utilizzo delle risorse su più tenant. Le device farm gestite offrono diversi tipi di dispositivi in modo da supportare hardware meno diffusi e di generazioni precedenti e da evitare l'impatto sulla sostenibilità dei clienti dovuti ad aggiornamenti dei dispositivi non necessari.

Risorse

Consulta le seguenti risorse per ulteriori informazioni sulle best practice per la sostenibilità.

Whitepaper

- Pilastro della sostenibilità

Video

- The Climate Pledge

Il processo di revisione

La revisione delle architetture va eseguita in modo coerente, con un approccio che non colpevolizza nessuno, ma che incoraggia ad approfondire gli argomenti. Dovrebbe essere un processo leggero (di ore, non di giorni) più simile a una conversazione che non a un audit. Lo scopo della revisione di un'architettura è identificare dei problemi critici da affrontare o aree di miglioramento. Il risultato della revisione è un insieme di azioni volte a migliorare l'esperienza di utilizzo del carico di lavoro del cliente.

Come discusso nella sezione "Architettura", ogni membro del team deve prendersi la responsabilità della qualità della sua architettura. Consigliamo che i membri del team che hanno sviluppato l'architettura usino il Framework Well-Architected per eseguire costantemente la revisione della loro architettura, piuttosto che fare una riunione di revisione formale. Un approccio quasi continuo permette ai membri del team di aggiornare le risposte man mano che l'architettura evolve e migliorare l'architettura di pari passo alle funzionalità.

Il AWS Well-Architected Framework è allineato al modo in cui esamina sistemi e servizi AWS internamente. Si basa su una serie di principi di progettazione che influenzano l'approccio architettonico e su domande volte a verificare che le persone non trascurino le aree spesso presenti in Root Cause Analysis (). RCA Ogni volta che si verifica un problema significativo con un sistema, un AWS servizio o un cliente interno, lo esaminiamo RCA per vedere se possiamo migliorare i processi di revisione che utilizziamo.

Le revisioni vanno applicate nelle tappe fondamentali del ciclo di vita del prodotto, nelle prime fasi di progettazione per evitare porte a un senso difficili da cambiare e prima della data di lancio. Molte decisioni sono porte reversibili a doppio senso e possono utilizzare un processo leggero. Le porte a un senso sono difficili o impossibili da invertire e richiedono ulteriori ispezioni prima della loro realizzazione. Una volta entrato in produzione, il carico di lavoro continuerà ad evolversi man mano che si aggiungono nuove funzionalità e si modificano le implementazioni tecnologiche. L'architettura del carico di lavoro cambia nel tempo. Devi seguire le best practice di igiene informatica per interrompere il degrado delle caratteristiche man mano che fai evolvere l'architettura. Man mano che l'architettura cambia, dovresti seguire un insieme di processi di igiene informatica tra cui la revisione Well-Architected.

Se vuoi utilizzare la revisione come snapshot una tantum o misura indipendente, dovrai verificare che alla conversazione partecipino tutte le persone appropriate. Spesso ci rendiamo conto che le revisioni sono il primo momento in cui il team comprende per davvero quello che ha implementato.

Un approccio che funziona bene per la revisione dei carichi di lavoro di un altro team consiste in una serie di conversazioni informali sull'architettura in cui ottenere le risposte alla maggior parte delle domande. Quindi puoi fare una o due riunioni di follow up in cui puoi fare chiarezza o approfondire le aree ambigue e il rischio percepito.

Ecco alcuni elementi suggeriti per le tue riunioni:

- Una sala riunioni con una lavagna
- Le stampe di tutti i grafici o delle note di progettazione
- Elenco di azioni a cui è necessaria una out-of-band ricerca per rispondere (ad esempio, «abbiamo attivato la crittografia o no?»)

Dopo avere completato la revisione, dovresti avere un elenco di problemi a cui assegnare delle priorità sulla base del contesto aziendale. Dovrai anche tenere conto dell'impatto di tali problemi sul day-to-day lavoro del tuo team. Se affronti questi problemi in anticipo puoi liberare del tempo per lavorare sulla creazione di valore aziendale anziché dedicarlo a risolvere i problemi ricorrenti. Man mano che affronti i problemi, puoi aggiornare la revisione per vedere in che modo l'architettura sta migliorando.

Il valore di una revisione è evidente dopo averne eseguita una, ma all'inizio un nuovo team potrebbe essere contrario. Ecco alcune obiezioni da gestire per istruire il team sui vantaggi di una revisione:

- "Siamo troppo occupati!" Spesso si sente questa frase quando il team si sta preparando a un grande lancio.
  - Se ti stai preparando per un grande lancio, desidererai che tutto vada bene. La revisione ti permetterà di individuare qualsiasi problema che potresti esserti perso.
  - Ti raccomandiamo di eseguire le revisioni all'inizio del ciclo di vita del prodotto per scoprire i rischi e sviluppare un piano di mitigazione allineato con la roadmap delle funzionalità.
- "Non abbiamo tempo per fare nulla per i risultati!" Spesso questo viene detto quando c'è un evento fisso, come il Super Bowl, di cui si sta occupando il team.
  - Questi eventi non possono essere spostati. Vuoi davvero affrontare l'evento senza conoscere i rischi della tua architettura? Anche se non ti occupi di tutti i problemi in questione, puoi comunque disporre di playbook per affrontarli se si dovessero presentare.
- "Non vogliamo che altri scoprano i segreti della nostra implementazione di soluzioni!"
  - Se poni le domande del Framework Well-Architected, il team noterà che nessuna di esse rivela informazioni proprietarie commerciali o tecniche.

Eseguendo più revisioni con i team della tua organizzazione, potresti identificare delle aree tematiche. Ad esempio, potresti notare che un gruppo di team ha gruppi di problemi in un pilastro o un argomento specifico. Puoi gestire tutte le tue revisioni in modo olistico e identificare tutti i meccanismi, la formazione o le riunioni con gli ingegneri responsabili che possono aiutare a risolvere i problemi tematici.

Conclusioni

Il AWS Well-Architected Framework fornisce le migliori pratiche architettoniche attraverso i sei pilastri per la progettazione e la gestione di sistemi affidabili, sicuri, efficienti, convenienti e sostenibili nel cloud. Il Framework fornisce un insieme di domande che ti permettono di eseguire la revisione di un'architettura esistente o proposta. Fornisce inoltre una serie di AWS best practice per ogni pilastro. L'utilizzo del Framework nella tua architettura ti aiuta a produrre sistemi stabili ed efficienti, che ti permettono di concentrarti sui tuoi requisiti funzionali.

Collaboratori

Le seguenti persone e organizzazioni hanno contribuito a questo documento:

- Brian Carlson, Operations Lead Well-Architected, Amazon Web Services
- Ben Potter, Security Lead Well-Architected, Amazon Web Services
- Seth Eliot, Reliability Lead Well-Architected, Amazon Web Services
- Eric Pullen, Sr. Solutions Architect, Amazon Web Services
- Rodney Lester, Principal Solutions Architect, Amazon Web Services
- Jon Steele, Sr. Technical Account Manager, Amazon Web Services
- Max Ramsay, Principal Security Solutions Architect, Amazon Web Services
- Callum Hughes, Solutions Architect, Amazon Web Services
- Ben Mergen, Senior Cost Lead Solutions Architect, Amazon Web Services
- Chris Kozlowski, Senior Specialist Technical Account Manager, Enterprise Support, Amazon Web Services
- Alex Livingstone, Principal Specialist Solutions Architect, Cloud Operations, Amazon Web Services
- Paul Moran, Principal Technologist, Enterprise Support, Amazon Web Services
- Peter Mullen, Advisory Consultant, Professional Services, Amazon Web Services
- Chris Pates, Senior Specialist Technical Account Manager, Enterprise Support, Amazon Web Services
- Arvind Raghunathan, Principal Specialist Technical Account Manager, Enterprise Support, Amazon Web Services
- Sam Mokhtari, Senior Efficiency Lead Solutions Architect, Amazon Web Services

Approfondimenti

AWS Architecture Center

Conformità di sicurezza nel cloud AWS

AWS Programma Well-Architected Partner

AWS Well-Architected Tool

AWS Homepage Well-Architected

Whitepaper sul pilastro dell'eccellenza operativa

Whitepaper sul pilastro della sicurezza

Whitepaper sul pilastro dell'affidabilità

Whitepaper sul pilastro dell'efficienza delle prestazioni

Whitepaper sul pilastro dell'ottimizzazione dei costi

Whitepaper sul pilastro della sostenibilità

Amazon Builders' Library

Revisioni del documento

Per ricevere notifiche sugli aggiornamenti di questo white paper, iscriviti al feed. RSS

Modifica	Descrizione	Data
Linee guida sulle best practice aggiornate	In tutti i pilastri sono stati apportati aggiornamenti su larga scala in merito alle best practice. Sono state predisposte best practice correlate a sicurezza e costi.	27 giugno 2024
Aggiornamento principale	Principali aggiornamenti dei pilastri.	3 ottobre 2023
Aggiornamenti per il nuovo framework	Best practice aggiornate con prontuario e nuove best practice aggiunte. Nuove domande aggiunte sui pilastri di sicurezza e di ottimizzazione dei costi.	10 aprile 2023
Aggiornamento secondario	Aggiunta della definizione di livello di impegno e aggiornamento delle best practice nell'appendice.	20 ottobre 2022
Aggiornamento del whitepaper	Aggiunta del pilastro della sostenibilità e collegamenti aggiornati.	2 dicembre 2021
Aggiornamento principale	Aggiunta al framework del pilastro della sostenibilità.	20 novembre 2021
Aggiornamento secondario	Rimozione del linguaggio non inclusivo.	22 aprile 2021

Aggiornamento secondario	Correzione di diversi collegamenti.	10 marzo 2021
Aggiornamento secondario	Modifiche editoriali di minore entità in varie parti del documento.	15 luglio 2020
Aggiornamenti per il nuovo framework	Revisione e riscrittura della maggior parte delle domande e delle risposte.	8 luglio 2020
Aggiornamento del whitepaper	Aggiunta di AWS Well-Arch itected Tool, collegamenti a AWS Well-Architected Labs AWS e Well-Architected Partners, correzioni minori per abilitare la versione multilingue del framework.	1° luglio 2019
Aggiornamento del whitepaper	Revisione e riscrittura di molte domande e risposte per garantire che le domande si concentrino su un argomento alla volta. Per questo motivo, alcune delle domande precedenti sono state divise in più domande. Aggiunta di termini comuni alle definizioni (carichi di lavoro, component i, ecc.). Presentazione delle domande modificata per includere il testo descrittivo.	1° novembre 2018
Aggiornamento del whitepaper	Aggiornamenti volti a semplific are il testo delle domande e a migliorare la leggibilità.	1° giugno 2018

Aggiornamento del whitepaper	Eccellenza operativa spostata all'inizio della sezione sui pilastri e riscritta in modo che inquadri gli altri pilastri. Sono stati aggiornati altri pilastri per riflettere l'evoluzione di. AWS	1° novembre 2017
Aggiornamento del whitepaper	Framework aggiornato per includere i pilastri dell'ecce llenza operativa; altri pilastri rivisti e aggiornati per ridurre la duplicazione e incorporare le nozioni apprese grazie alle revisioni eseguite con migliaia di clienti.	1° novembre 2016
Aggiornamenti minori	È stata aggiornata l'Appendice con le informazioni correnti di Amazon CloudWatch Logs.	1° novembre 2015
Pubblicazione iniziale	AWS Pubblicato il Well-Arch itected Framework.	1° ottobre 2015

> **Note**
>
> Per sottoscrivere RSS gli aggiornamenti, devi avere un RSS plugin abilitato per il browser che stai utilizzando.

Versione del Framework

- 03/10/2023 (attuale)
- 10/04/2023
- 2022-03-31

Appendice: domande e best practice

Questa appendice riassume tutte le domande e le best practice nel Framework AWS Well-Architected.

Pilastri

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

Eccellenza operativa

Il pilastro dell'eccellenza operativa include la capacità di supportare lo sviluppo ed eseguire carichi di lavoro in modo efficace, ottenere informazioni approfondite sulle operazioni e migliorare continuamente i processi e le procedure di supporto per offrire valore commerciale. Puoi trovare linee guida prescrittive sull'implementazione nel whitepaper sul pilastro dell'eccellenza operativa.

Aree delle best practice

- Organizzazione
- Preparazione
- Gestione
- Evoluzione

Organizzazione

Questions

- OPS1. Come stabilisci quali sono le tue priorità?
- OPS2. Come strutturare la tua organizzazione per supportare i risultati aziendali?
- OPS3. In che modo la cultura aziendale supporta i risultati aziendali?

OPS1. Come stabilisci quali sono le tue priorità?

È necessario che ognuno comprenda il proprio ruolo per rendere possibile il successo aziendale. Devi disporre di obiettivi comuni al fine di stabilire le priorità per le risorse. Ciò massimizzerà i risultati dei tuoi sforzi.

Best practice

- OPS01-BP01 Valuta le esigenze dei clienti
- OPS01-BP02 Valuta le esigenze interne dei clienti
- OPS01-BP03 Valuta i requisiti di governance
- OPS01-BP04 Valuta i requisiti di conformità
- OPS01-BP05 Valuta il panorama delle minacce
- OPS01-BP06 Valuta i compromessi gestendo vantaggi e rischi

OPS01-BP01 Valuta le esigenze dei clienti

Coinvolgi le principali parti interessate, compresi i team aziendali, di sviluppo e operativi, per determinare dove concentrare gli sforzi in base alle esigenze dei clienti esterni. Avrai così una conoscenza approfondita del supporto operativo necessario per raggiungere i risultati aziendali desiderati.

Risultato desiderato:

- Lavori a ritroso partendo dalle esigenze dei clienti.
- Comprendi in che modo le procedure operative supportano i risultati e gli obiettivi aziendali.
- Coinvolgi tutte le parti interessate.
- Disponi di meccanismi per soddisfare le esigenze dei clienti.

Anti-pattern comuni:

- Hai deciso di non fornire il servizio clienti al di fuori dell'orario lavorativo di base, ma non hai esaminato i dati cronologici riguardanti le richieste di supporto. Non sai se questo determinerà un impatto sui tuoi clienti.
- Stai sviluppando una nuova funzionalità, ma non hai coinvolto i clienti per capire se è desiderata ed eventualmente in quale forma; inoltre non hai condotto attività di sperimentazione per convalidarne la necessità e il metodo di distribuzione.

Vantaggi dell'adozione di questa best practice: i clienti le cui esigenze sono soddisfatte hanno maggiori probabilità di rimanere clienti. Valutando e comprendendo le esigenze dei clienti esterni sarà possibile organizzare le attività in base alle priorità e offrire valore aggiunto.

Livello di rischio associato se questa best practice non fosse adottata: elevato

Guida all'implementazione

Comprendi le esigenze aziendali: il successo dell'azienda nasce dalla condivisione di obiettivi e conoscenze tra le parti interessate, compresi i team aziendali, di sviluppo e operativi.

Esamina gli obiettivi aziendali, le esigenze e le priorità dei clienti esterni: coinvolgi le principali parti interessate, compresi i team aziendali, di sviluppo e operativi, per discutere obiettivi, esigenze e priorità dei clienti esterni. In tal modo otterrai una conoscenza approfondita del supporto operativo necessario per raggiungere i risultati aziendali e del cliente.

Stabilisci una comprensione condivisa: stabilisci una comprensione condivisa delle funzioni aziendali del carico di lavoro, dei ruoli di ciascuno dei team nel gestire il carico di lavoro e di come questi supportino i tuoi obiettivi aziendali condivisi tra clienti interni ed esterni.

Risorse

Best practice correlate:

- OPS11-BP03 Implementa cicli di feedback

OPS01-BP02 Valuta le esigenze interne dei clienti

Coinvolgi le principali parti interessate, compresi i team aziendali, di sviluppo e operativi, nel determinare dove concentrare le attività in base alle esigenze dei clienti interni. Questo ti garantirà una conoscenza approfondita del supporto operativo necessario per raggiungere i risultati aziendali.

Risultato desiderato:

- Utilizzi le priorità definite per concentrare le iniziative di miglioramento delle operazioni laddove avranno il maggiore impatto (ad esempio, sviluppare le competenze dei team, migliorare le prestazioni del carico di lavoro, ridurre i costi, automatizzare i runbook o potenziare il monitoraggio).
- Aggiorni le priorità al mutare delle esigenze.

Anti-pattern comuni:

- Per semplificare la gestione della rete hai deciso di modificare l'assegnazione degli indirizzi IP per i team di prodotto senza consultarli. Non conosci l'impatto che questo avrà sui tuoi team di prodotto.
- Stai implementando un nuovo strumento di sviluppo, ma non hai coinvolto i clienti interni per scoprire se è necessario o se è compatibile con le loro pratiche esistenti.
- Stai implementando un nuovo sistema di monitoraggio, ma non hai contattato i clienti interni per scoprire se hanno esigenze di monitoraggio o reporting da tenere in considerazione.

Vantaggi dell'adozione di questa best practice: valutando e comprendendo le esigenze dei clienti interni consente di organizzare le attività in base a priorità e offrire valore aggiunto.

Livello di rischio associato se questa best practice non fosse adottata: elevato

Guida all'implementazione

- Comprendi le esigenze aziendali: il successo dell'azienda nasce dalla condivisione di obiettivi e conoscenze tra le parti interessate, compresi i team aziendali, di sviluppo e operativi.
- Esamina gli obiettivi aziendali, le esigenze e le priorità dei clienti interni: coinvolgi le principali parti interessate, compresi i team aziendali, di sviluppo e operativi, per discutere obiettivi, esigenze e priorità dei clienti interni. In tal modo otterrai una conoscenza approfondita del supporto operativo necessario per raggiungere i risultati aziendali e del cliente.
- Stabilisci una comprensione condivisa: stabilisci una comprensione condivisa delle funzioni aziendali del carico di lavoro, dei ruoli di ciascuno dei team nel gestire il carico di lavoro e di come questi supportino gli obiettivi aziendali condivisi tra clienti interni ed esterni.

Risorse

Best practice correlate:

- OPS11-BP03 Implementa cicli di feedback

OPS01-BP03 Valuta i requisiti di governance

Con governance si intende l'insieme di policy, regole o framework che un'azienda usa per raggiungere i propri obiettivi. I requisiti di governance vengono generati all'intero dell'organizzazione. Possono influire sui tipi di tecnologia che scegli o sul modo in cui esegui il tuo carico di lavoro. Integra

i requisiti di governance della tua organizzazione nel tuo carico di lavoro. La conformità è la capacità di dimostrare che hai implementato i requisiti di governance.

Risultato desiderato:

- I requisiti di governance sono integrati nel progetto architetturale e nell'operatività del tuo carico di lavoro.
- Puoi dimostrare di aver seguito i requisiti di governance.
- I requisiti di governance vengono rivisti e aggiornati con regolarità.

Anti-pattern comuni:

- La tua azienda richiede che l'account root abbia l'autenticazione multi-fattore. Non sei riuscito a implementare questo requisito e l'account root è compromesso.
- Durante la progettazione del carico di lavoro hai scelto un tipo di istanza non approvata dal dipartimento IT. Non riesci ad avviare il tuo carico di lavoro e devi procedere a una nuova progettazione.
- Devi avere un piano di ripristino di emergenza. Non ne hai uno e il tuo carico di lavoro è vittima di un'interruzione prolungata.
- Il tuo team vuole usare nuove istanze, ma i requisiti di governance non sono stati aggiornati e pertanto non sono consentite.

Vantaggi dell'adozione di questa best practice:

- Rispettare i requisiti di governance permette di allineare il carico di lavoro a policy organizzative di più ampio respiro.
- I requisiti di governance si basano su standard e best practice di settore per la tua organizzazione.

Livello di rischio associato se questa best practice non fosse adottata: elevato

Guida all'implementazione

Identifica il requisito di governance collaborando con le parti interessate e le organizzazioni preposte. Includi i requisiti di governance nel tuo carico di lavoro. Dimostra di aver seguito i requisiti di governance.

Esempio del cliente

In AnyCompany Retail, il team operativo del cloud collabora con le parti interessate di tutta l'organizzazione per sviluppare requisiti di governance. Ad esempio, vietano l'SSHaccesso alle EC2 istanze Amazon. Se i team hanno necessità di accedere ai sistemi, devono usare AWS Systems Manager Session Manager. Il team operativo nell'ambiente cloud aggiorna con regolarità i requisiti di governance nel momento in cui vengono rilasciati nuovi servizi.

Passaggi dell'implementazione

1. Identifica le parti interessate per il tuo carico di lavoro, inclusi eventuali team centralizzati.
2. Collabora con le parti interessate per identificare i requisiti di governance.
3. Dopo aver generato un elenco, dai la priorità alle voci relative a migliorie e inizia a implementarle nel tuo carico di lavoro.
   a. Utilizza servizi come AWS Configcreare governance-as-code e convalidare il rispetto dei requisiti di governance.
   b. Utilizzando AWS Organizations, puoi avvalerti di policy di controllo dei servizi per l'implementazione dei requisiti di governance.
4. Fornisci la documentazione che convalida l'implementazione.

Livello di impegno per il piano di implementazione: medio L'implementazione di requisiti di governance mancanti può causare la rielaborazione del tuo carico di lavoro.

Risorse

Best practice correlate:

- OPS01-BP04 Valuta i requisiti di conformità: la conformità è come la governance, ma è esterna rispetto all'organizzazione.

Documenti correlati:

- AWS Guida all'ambiente cloud di gestione e governance
- Le migliori pratiche per le politiche di controllo dei AWS Organizations servizi in un ambiente con più account
- Governance in Cloud AWS: Il giusto equilibrio tra agilità e sicurezza
- Cosa sono la governance, il rischio e la conformità (GRC)?

Video correlati:

- AWS Gestione e governance: configurazione, conformità e audit - AWS Online Tech Talks
- AWS RE:InForce 2019: governance per l'era del cloud (-R1) DEM12
- AWS re:Invent 2020: raggiungi la conformità come codice utilizzando il codice AWS Config
- AWS re:Invent 2020: governance agile su AWS GovCloud (US)

Esempi correlati:

- AWS Config Esempi di Conformance Pack

Servizi correlati:

- AWS Config
- AWS Organizations - Politiche di controllo dei servizi

OPS01-BP04 Valuta i requisiti di conformità

I requisiti di conformità interna, di settore e normativa sono un fattore importante per la definizione delle priorità della tua organizzazione. L'assetto di conformità della tua azienda potrebbe impedirti di usare tecnologie specifiche o posizioni geografiche. Applica la due diligence in assenza di contesti di conformità esterni. Genera audit o report per convalidare la conformità.

Se comunichi all'esterno che il tuo prodotto è in linea con standard specifici di conformità, devi disporre di un processo interno in grado di garantire in modo costante la conformità. Esempi di standard di conformità includono PCIDSS, Fed e. RAMP HIPAA Gli standard di conformità applicabili vengono stabiliti in base a diversi fattori, come il tipo di dati che la soluzione archivia o trasmette e quali aree geografiche sono supportate dalla soluzione.

Risultato desiderato:

- Requisiti di conformità interni, di settore e normativi sono integrati nella selezione dell'architettura.
- Puoi verificare la conformità e generare report di audit.

Anti-pattern comuni:

- Parte del carico di lavoro rientra nel framework Payment Card Industry Data Security Standard (PCI-DSS), ma il carico di lavoro archivia i dati delle carte di credito in modo non crittografato.
- Architetti e sviluppatori software non conoscono il contesto di conformità che la tua organizzazione è tenuta a rispettare.
- L'audit annuale di Systems and Organizations Control (SOC2) Type II avrà luogo a breve e non sarà possibile verificare che i controlli siano in atto.

Vantaggi dell'adozione di questa best practice:

- Grazie alla valutazione e comprensione dei requisiti di conformità applicati al carico di lavoro, sarà possibile organizzare le attività in base a priorità e offrire valore aggiunto.
- Scegli le sedi e le tecnologie corrette, in linea con il tuo contesto di integrità.
- La progettazione del tuo carico di lavoro ai fini degli audit ti consente di dimostrare il rispetto del modello di conformità.

Livello di rischio associato se questa best practice non fosse adottata: elevato

Guida all'implementazione

L'implementazione di questa best practice significa integrare requisiti di conformità nel processo di progettazione dell'architettura. I membri del tuo team sono a conoscenza del contesto di conformità richiesto. Convalidi la conformità in linea con il contesto.

Esempio del cliente

AnyCompany AI dettaglio archivia i dati delle carte di credito dei clienti. Gli sviluppatori del team di archiviazione delle carte comprendono che devono rispettare il DSS framework PCI -. Hanno adottato misure per verificare che i dati della carta di credito siano archiviati e accessibili in modo sicuro in linea con il PCI - DSS framework. Ogni anno collaborano con il team di sicurezza per confermare la conformità.

Passaggi dell'implementazione

1. Collabora con i team di sicurezza e governance per stabilire le conformità interne, normative o di settore deve rispettare il tuo carico di lavoro. Integra gli standard di conformità nel tuo carico di lavoro.
   a. Convalida la conformità continua delle AWS risorse con servizi come e. AWS Compute OptimizerAWS Security Hub

2. Comunica ai membri del tuo team i requisiti di conformità, in modo che possano gestire e far evolvere il carico di lavoro in linea con essi. I requisiti di conformità devono essere inclusi nelle scelte tecnologiche e architetturali.
3. A seconda del contesto di conformità, potresti dover generare un report di audit o conformità. Collabora con la tua organizzazione per automatizzare il più possibile questo processo.
   a. Utilizza servizi come AWS Audit Manager per convalidare la conformità e generare report di audit.
   b. Puoi scaricare i documenti AWS di sicurezza e conformità con. AWS Artifact

Livello di impegno per il piano di implementazione: medio Implementare i requisiti di conformità può essere complesso. Generare report di audit o documenti di conformità aggiunge altre complessità.

Risorse

Best practice correlate:

- SEC01-BP03 Identificare e convalidare gli obiettivi di controllo - Gli obiettivi di controllo della sicurezza sono una parte importante della conformità generale.
- SEC01-BP06 Automatizza i test e la convalida dei controlli di sicurezza nelle pipeline: come parte delle tue pipeline, convalida i controlli di sicurezza. Puoi anche generare la documentazione di conformità per le nuove modifiche.
- SEC07-BP02 Definizione dei controlli di protezione dei dati - Molti framework di conformità si basano su politiche di gestione e archiviazione dei dati.
- SEC10-BP03 Prepara le funzionalità forensi: a volte le funzionalità forensi possono essere utilizzate per verificare la conformità.

Documenti correlati:

- AWS Centro di conformità
- AWS Risorse per la conformità
- AWS White paper su rischi e conformità
- AWS Modello di responsabilità condivisa
- AWS servizi contemplati dai programmi di conformità

Video correlati:

- AWS re:Invent 2020: raggiungi la conformità come codice utilizzando AWS Compute Optimizer
- AWS re:Invent 2021 - Conformità, garanzia e audit del cloud
- AWS Summit ATL 2022 - Implementazione della conformità, della garanzia e del controllo su (02) AWS COP2

Esempi correlati:

- PCIDSSe le migliori AWS pratiche di sicurezza fondamentali su AWS

Servizi correlati:

- AWS Artifact
- AWS Audit Manager
- AWS Compute Optimizer
- AWS Security Hub

OPS01-BP05 Valuta il panorama delle minacce

Valuta le minacce per l'azienda (ad esempio, concorrenza, rischi e responsabilità aziendali, rischi operativi e minacce per la sicurezza delle informazioni) e conserva le informazioni aggiornate in un registro dei rischi. Quando stabilisci dove concentrare gli sforzi, tieni in considerazione l'impatto dei rischi.

Il Framework Well-Architected enfatizza formazione, misurazione e miglioramento. Fornisce un approccio coerente per valutare le architetture e implementare progetti scalabili nel tempo. AWS fornisce l'assistenza necessaria AWS Well-Architected Toolper rivedere l'approccio prima dello sviluppo, lo stato dei carichi di lavoro prima della produzione e lo stato dei carichi di lavoro in produzione. Puoi confrontarli con le migliori pratiche AWS architettoniche più recenti, monitorare lo stato generale dei carichi di lavoro e ottenere informazioni sui potenziali rischi.

AWS i clienti hanno diritto a una revisione guidata Well-Architected dei loro carichi di lavoro mission-critical per misurare le loro architetture rispetto alle migliori pratiche. AWS I clienti del supporto Enterprise possono usufruire di una revisione delle operazioni, ideata per agevolare l'identificazione di lacune nell'approccio da loro utilizzato nel cloud.

Il coinvolgimento trasversale dei team per tali controlli aiuta a comprendere a livello comune i carichi di lavoro e il contributo dei ruoli del team al successo. Le esigenze identificate nel corso dell'analisi possono aiutarti a definire le priorità.

AWS Trusted Advisor è uno strumento che fornisce l'accesso a una serie di controlli di base che propongono ottimizzazioni utili per la definizione delle tue priorità. I clienti del supporto Business ed Enterprise hanno accesso a ulteriori controlli a livello di sicurezza, affidabilità, prestazioni e ottimizzazione dei costi, utili per definire le loro priorità.

Risultato desiderato:

- Esamini e agisci regolarmente in base a Well-Architected Trusted Advisor e ai risultati
- Sei a conoscenza dello stato delle patch più recenti dei servizi.
- Comprendi il rischio e l'impatto delle minacce note e intervieni di conseguenza.
- Implementi le mitigazioni necessarie.
- Comunichi azioni e contesto.

Anti-pattern comuni:

- Utilizzo della versione precedente di una libreria software nel tuo prodotto. Mancata conoscenza di aggiornamenti di sicurezza alla libreria per problemi che potrebbero avere un impatto imprevisto sul carico di lavoro.
- Rilascio da parte di un tuo concorrente di una versione del proprio prodotto che risolve i reclami di molti dei tuoi clienti relativi al tuo prodotto. Non hai dato priorità alla risoluzione di questi problemi noti.
- Perseguimento da parte delle autorità di regolamentazione di aziende come la tua, non conformi ai requisiti di conformità alla normativa legale. Mancata assegnazione della priorità ai requisiti di conformità in sospeso.

Vantaggi dell'adozione di questa best practice: identifichi e comprendi le minacce per la tua organizzazione e il tuo carico di lavoro ti consentono di determinare quali minacce affrontare, la loro priorità e le risorse necessarie per farlo.

Livello di rischio associato se questa best practice non fosse adottata: medio

Guida all'implementazione

- Valuta il panorama delle minacce: valute le minacce per l'azienda (ad esempio, concorrenza, rischi e responsabilità aziendali, rischi operativi e minacce alla sicurezza delle informazioni), in modo da poterne includere l'impatto nel determinare dove concentrare le attività.
  - Bollettini sulla sicurezza AWS aggiornati
  - AWS Trusted Advisor
- Mantieni un modello delle minacce: definisci e mantieni un modello delle minacce che identifichi potenziali minacce, mitigazioni pianificate e predisposte e la relativa priorità. Esamina la probabilità che le minacce si manifestino come incidenti, il costo del ripristino dagli incidenti, il danno previsto causato e il costo per prevenire tali incidenti. Modifica le priorità man mano che i contenuti del modello di minaccia cambiano.

Risorse

Best practice correlate:

- SEC01-BP07 Identifica le minacce e dai priorità alle mitigazioni utilizzando un modello di minaccia

Documenti correlati:

- Conformità di Cloud AWS
- Bollettini sulla sicurezza AWS aggiornati
- AWS Trusted Advisor

Video correlati:

- AWS re:Inforce 2023 - A tool to help improve your threat modeling

OPS01-BP06 Valuta i compromessi gestendo vantaggi e rischi

Gli interessi contrastanti di più parti possono complicare l'assegnazione delle priorità a impegni, sviluppo delle capacità e conseguimento di risultati in linea con le strategie aziendali. Ad esempio, potrebbe esserti chiesto di accelerare l'introduzione speed-to-market di nuove funzionalità rispetto all'ottimizzazione dei costi dell'infrastruttura IT. Questa richiesta può mettere due parti interessate in

conflitto reciproco. In queste situazioni, le decisioni devono essere prese da un'autorità superiore che risolve il conflitto. I dati sono necessari per rimuovere l'aspetto emotivo dal processo decisionale.

La stessa sfida può verificarsi a livello strategico. Ad esempio, la scelta tra l'utilizzo di tecnologie di database relazionali o non relazionali può avere un impatto significativo sul funzionamento di un'applicazione. È fondamentale comprendere i risultati prevedibili delle varie scelte.

AWS può aiutarvi a istruire i vostri team in merito AWS ai suoi servizi per aumentare la loro comprensione di come le loro scelte possono avere un impatto sul carico di lavoro. Per istruire i tuoi team, utilizza le risorse fornite da AWS Support (Centro conoscenze AWS, AWS Discussion Forums e AWS Support Center) e la documentazione AWS. Per ulteriori domande, contatta. AWS Support

AWS condivide inoltre best practice e modelli operativi in The Amazon Builders' Library. Un'ampia varietà di altre informazioni utili è disponibile attraverso il AWS blog e il podcast ufficiale AWS.

Risultato desiderato: presenza di un framework di governance decisionale definito in modo chiaro per semplificare le decisioni importanti a tutti i livelli all'interno dell'organizzazione di distribuzione del cloud. Questo framework include funzionalità come registro dei rischi, ruoli definiti autorizzati a prendere decisioni e modelli prestabiliti per ogni livello di decisione adottabile. Il framework definisce in anticipo le modalità di risoluzione dei conflitti, quali dati vanno presentati e come viene stabilita la priorità delle opzioni, in modo che una volta prese le decisioni sia subito possibile lavorare per applicarle. Il framework del processo decisionale include un approccio standardizzato alla revisione e alla valutazione di vantaggi e rischi di ogni decisione per comprenderne i compromessi. Ciò può comprendere fattori esterni, come l'aderenza ai requisiti di conformità normativa.

Anti-pattern comuni:

- I vostri investitori richiedono che dimostriate la conformità agli standard di sicurezza dei dati del settore delle carte di pagamento (PCIDSS). Non prendi in considerazione i compromessi tra soddisfare la loro richiesta e continuare con le attività di sviluppo già in corso. Al contrario, prosegui con il lavoro di sviluppo senza dimostrare la conformità. Gli investitori interrompono il supporto all'azienda a causa dei dubbi relativi alla sicurezza della piattaforma e dei loro investimenti.
- Decisione di includere una libreria che uno dei tuoi sviluppatori ha trovato su Internet. Non hai valutato i rischi derivanti dall'adozione di questa libreria da un'origine sconosciuta e non sai se contiene vulnerabilità o codice dannoso.
- Giustificazione aziendale originale per la migrazione basata sulla modernizzazione del 60% dei carichi di lavoro delle applicazioni. Tuttavia, a causa di difficoltà tecniche, è stata presa la decisione di modernizzare solo il 20%, con una riduzione dei vantaggi pianificati a lungo termine, un maggiore impegno operativo dei team dell'infrastruttura per supportare manualmente i sistemi

legacy e un accresciuto affidamento sullo sviluppo di nuove competenze nei team dell'infrastruttura che non avevano pianificato questo cambiamento.

Vantaggi dell'adozione di questa best practice: allineamento e supporto completi delle priorità aziendali a livello gestionale, comprensione dei rischi legati al raggiungimento del successo, decisioni informate e azioni opportune quando i rischi costituiscono un ostacolo per le possibilità di successo. Comprendere implicazioni e conseguenze delle tue decisioni ti aiuta a stabilire le priorità delle opzioni, oltre a ottenere l'accordo dei leader più rapidamente, fornendo risultati aziendali migliori. L'identificazione dei benefici disponibili delle tue scelte e la consapevolezza dei rischi per la tua organizzazione ti aiutano a prendere decisioni basate sui dati, piuttosto che affidarti agli aneddoti.

Livello di rischio associato se questa best practice non fosse adottata: medio

Guida all'implementazione

La gestione di vantaggi e rischi va definita da un organo direttivo che stabilisca i requisiti del processo decisionale chiave. Le decisioni devono essere prese e la priorità va assegnata in base ai vantaggi derivanti per l'organizzazione, con una comprensione dei rischi connessi. L'accuratezza delle informazioni è fondamentale quando si prendono le decisioni organizzative, che devono basarsi su misurazioni affidabili ed essere definite secondo le procedure comuni del settore per l'analisi costi-benefici. Per prendere questo tipo di decisioni, occorre trovare un equilibrio tra l'autorità centralizzata e quella decentralizzata. Esiste sempre un compromesso ed è importante capire l'impatto di ogni scelta sulle strategie definite e sui risultati aziendali desiderati.

Passaggi dell'implementazione

1. Formalizza le procedure di misurazione dei vantaggi in un framework olistico di governance del cloud.
   a. Bilancia il controllo centrale del processo decisionale con l'autorità decentralizzata per alcune decisioni.
   b. Riconosci che i gravosi processi decisionali imposti per ogni decisione possono rallentare le operazioni.
   c. Incorpora nel processo decisionale fattori esterni, come i requisiti di conformità.
2. Stabilisci un framework del processo decisionale concordato per vari livelli di decisioni, che includa chi è tenuto a prendere le decisioni soggette a conflitti di interessi.
   a. Centralizza le decisioni definitive che potrebbero essere irreversibili.
   b. Consenti ai leader dell'organizzazione di livello inferiore di prendere decisioni reversibili.

3. Comprendi e gestisci i vantaggi e i rischi. Bilancia i vantaggi delle decisioni rispetto ai rischi connessi.
   a. Identificazione dei vantaggi: identifica i vantaggi in base a obiettivi aziendali, esigenze e priorità, Gli esempi includono l'impatto sui business case time-to-market, la sicurezza, l'affidabilità, le prestazioni e i costi.
   b. Identificazione dei rischi: identifica i rischi in base a obiettivi aziendali, esigenze e priorità, Gli esempi includono sicurezza time-to-market, affidabilità, prestazioni e costi.
   c. Valutazione dei vantaggi rispetto ai rischi e decisioni informate: determina l'impatto di vantaggi e rischi in base a obiettivi, esigenze e priorità delle principali parti interessate, inclusi business, sviluppo e operazioni. Valuta il valore del vantaggio rispetto alla probabilità di concretizzazione del rischio e al costo del suo impatto. Ad esempio, enfatizzare speed-to-market l'affidabilità potrebbe fornire un vantaggio competitivo. Tuttavia, ciò potrebbe causare tempi di attività ridotti in presenza di problemi di affidabilità.
4. Applica in modo programmatico le decisioni chiave che automatizzano l'aderenza ai requisiti di conformità.
5. Sfrutta strutture e funzionalità di settore note, come Value Stream AnalysisLEAN, per basare le prestazioni e le metriche aziendali allo stato attuale e definire le iterazioni dei progressi verso il miglioramento di tali metriche.

Livello di impegno per il piano di implementazione: medio-alto

Risorse

Best practice correlate:

- OPS01-BP05 Valuta il panorama delle minacce

Documenti correlati:

- Elementi della cultura del Giorno 1 di Amazon | Adotta decisioni di alta qualità e ad alta velocità
- Governance del cloud
- Management & Governance Cloud Environment
- Governance in the Cloud and in the Digital Age: Parts One & Two

Video correlati:

- Podcast | Jeff Bezos | On how to make decisions

Esempi correlati:

- Prendi decisioni informate utilizzando i dati (The Sagas) DevOps
- Utilizzo della mappatura del flusso di valore dello sviluppo per identificare i vincoli ai risultati DevOps

OPS2. Come strutturare la tua organizzazione per supportare i risultati aziendali?

I tuoi team devono comprendere quale contributo offrono nel raggiungimento dei risultati aziendali. I team devono avere obiettivi condivisi e comprendere il proprio ruolo nel successo degli altri team. Comprendere la responsabilità, la proprietà, il modo in cui vengono prese le decisioni e chi ha l'autorità decisionale aiuterà a concentrare gli sforzi e a ottimizzare i contributi dei team.

Best practice

- OPS02-BP01 Le risorse hanno identificato i proprietari
- OPS02-BP02 I processi e le procedure hanno identificato i proprietari
- OPS02-BP03 Le attività operative hanno identificato i proprietari responsabili delle loro prestazioni
- OPS02-BP04 Esistono meccanismi per gestire le responsabilità e la proprietà
- OPS02-BP05 Esistono meccanismi per richiedere aggiunte, modifiche ed eccezioni
- OPS02-BP06 Le responsabilità tra i team sono predefinite o negoziate

OPS02-BP01 Le risorse hanno identificato i proprietari

Le risorse per il tuo carico di lavoro devono disporre di proprietari identificati per il controllo delle modifiche, la risoluzione dei problemi e altre funzioni. I proprietari sono assegnati a carichi di lavoro, account, infrastrutture, piattaforme e applicazioni. La registrazione della proprietà avviene tramite strumenti come un registro centrale o metadati collegati alle risorse. Il valore aziendale dei componenti è alla base dei processi e delle procedure applicate.

Risultato desiderato:

- Le risorse presentano proprietari identificati tramite i metadati o un registro centrale.
- I membri del team possono identificare chi è il proprietario delle risorse.
- Gli account hanno un solo proprietario, laddove possibile.

Anti-pattern comuni:

- I tuoi contatti alternativi non sono popolati. Account AWS
- Risorse prive di tag che identificano i team proprietari.
- Hai una ITSM coda senza una mappatura delle email.
- Due team con entrambi la proprietà di una parte critica dell'infrastruttura.

Vantaggi dell'adozione di questa best practice:

- Il controllo delle modifiche per le risorse è immediato con la proprietà assegnata.
- Puoi coinvolgere i proprietari corretti quando risolvi i problemi.

Livello di rischio associato se questa best practice non fosse adottata: elevato

Guida all'implementazione

Definisci qual è il significato della proprietà per i casi d'uso delle risorse nel tuo ambiente. Proprietà significa supervisionare le modifiche alla risorsa, supportare la risorsa durante la risoluzione dei problemi o essere finanziariamente affidabile. Specifica e registra i proprietari delle risorse, con nome, informazioni di contatto, organizzazione e team.

Esempio del cliente

AnyCompany La vendita al dettaglio definisce la proprietà come il team o l'individuo responsabile delle modifiche e del supporto alle risorse. Sfruttano AWS Organizations per gestire le proprie Account AWS. Contatti alternativi degli account sono configurati con caselle di posta di gruppo. Ogni ITSM coda è associata a un alias e-mail. I tag identificano chi possiede le risorse. AWS Per altre piattaforme e infrastrutture, è presente una pagina wiki che identifica proprietà e informazioni di contatto.

Passaggi dell'implementazione

1. Inizia definendo la proprietà dell'organizzazione. La proprietà può significare essere proprietari del rischio collegato alla risorsa, delle modifiche alla risorsa o supportare la stessa durante la risoluzione dei problemi. Proprietà può anche significare proprietà amministrativa o finanziaria della risorsa.
2. Usa AWS Organizations per gestire gli account. Puoi gestire a livello centrale i contatti alternativi per gli account.

   a. Se usi indirizzi e-mail e numeri di telefono aziendali come informazioni di contatto, puoi accedervi anche se le persone a cui appartengono non fanno più parte dell'organizzazione. Ad esempio, crea elenchi di distribuzione delle e-mail separati per fatturazione, operazioni e sicurezza e configurali come contatti per Fatturazione, Sicurezza e Operazioni in ogni Account AWS attivo. Più persone riceveranno AWS notifiche e saranno in grado di rispondere, anche se qualcuno è in vacanza, cambia ruolo o lascia l'azienda.
   b. Se un account non è gestito da AWS Organizations, i contatti alternativi dell'account aiutano AWS a contattare il personale opportuno, se necessario. Configura i contatti alternativi dell'account per indirizzare le persone a un gruppo invece che a un individuo.
3. Utilizza i tag per identificare i proprietari AWS delle risorse. Puoi specificare i proprietari e le loro informazioni di contatto in tag separati.
   a. Puoi utilizzare le regole di AWS Config per far sì che le risorse presentino i tag di proprietà richiesti.
   b. Per una guida approfondita su come creare una strategia di tagging per la tua organizzazione, consulta il whitepaper AWS Tagging Best Practices.
4. Usa Amazon Q Business, un assistente conversazionale che utilizza l'IA generativa per migliorare la produttività della forza lavoro, rispondere a domande e completare attività in base alle informazioni presenti nei sistemi aziendali.
   a. Collega Amazon Q Business all'origine dati della tua azienda. Amazon Q Business offre connettori predefiniti per oltre 40 fonti di dati supportate, tra cui Amazon Simple Storage Service (Amazon S3), SharePoint Microsoft, Salesforce e Atlassian Confluence. Per ulteriori informazioni, consulta Connettori di Amazon Q Business.
5. Per altre risorse, piattaforme e infrastrutture, crea la documentazione che stabilisce la proprietà. Tutti i membri del team devono poter accedere a queste informazioni.

Livello di impegno per il piano di implementazione: basso Sfrutta le informazioni di contatto e i tag dell'account per assegnare la proprietà delle risorse. AWS Per altre risorse puoi usare qualcosa di semplice come una tabella in un wiki per registrare la proprietà e le informazioni di contatto, oppure usare ITSM uno strumento per mappare la proprietà.

Risorse

Best practice correlate:

- OPS02-BP02 I processi e le procedure hanno identificato i proprietari
- OPS02-BP04 Esistono meccanismi per gestire le responsabilità e la proprietà

Documenti correlati:

- AWS Account Management - Updating contact information
- AWS Organizations - Aggiornamento dei contatti alternativi all'interno dell'organizzazione
- Whitepaper AWS Tagging Best Practices
- Crea app di intelligenza artificiale generativa aziendali private e sicure con Amazon Q Business e AWS IAM Identity Center
- Amazon Q Business, now generally available, helps boost workforce productivity with generative AI
- Cloud AWS Blog Operations & Migrations - Implementazione di controlli di tagging automatizzati e centralizzati con e AWS ConfigAWS Organizations
- AWS Blog sulla sicurezza - Estendi i tuoi hook di pre-commit con AWS CloudFormation Guard
- AWS DevOps Blog - Integrazione AWS CloudFormation Guard nelle pipeline CI/CD

Workshop correlati:

- Workshop AWS - Tagging

Esempi correlati:

- Regole di AWS Config - Amazon EC2 con tag obbligatori e valori validi

Servizi correlati:

- Regole di AWS Config - tag obbligatori
- AWS Organizations

OPS02-BP02 I processi e le procedure hanno identificato i proprietari

È utile sapere chi ha la proprietà della definizione di singoli processi e procedure, poiché tali processi e procedure specifici vengono utilizzati e perché tale proprietà esiste. Comprendere i motivi per cui vengono utilizzati processi e procedure specifici aiuta a identificare le opportunità di miglioramento.

Risultato desiderato: la tua organizzazione dispone di una serie di processi e procedure per le attività operative ben definiti e gestiti. L'archiviazione di processi e procedure avviene in una posizione centrale e questi sono a disposizione dei membri del team. I processi e le procedure vengono

aggiornati di frequente attraverso l'assegnazione chiara della proprietà. Ove possibile, script, modelli e documenti di automazione vengono implementati come codice.

Anti-pattern comuni:

- Mancata documentazione dei processi. È possibile la presenza di script frammentati su workstation degli operatori isolate.
- Conoscenza relativa all'uso degli script nelle mani di pochi individui oppure l'acquisizione avviene in modo informale come conoscenza di team.
- Necessità di aggiornare un processo legacy, ma manca chiarezza circa la proprietà dell'aggiornamento e l'autore originale non fa più parte dell'organizzazione.
- Non è possibile individuare processi e script, quindi non sono immediatamente disponibili quando necessario (ad esempio, in risposta a un incidente).

Vantaggi dell'adozione di questa best practice:

- Processi e procedure incentivano l'impegno nella gestione dei carichi di lavoro.
- I nuovi membri del team diventano efficienti in modo più rapido.
- Riduzione dei tempi di mitigazione degli incidenti.
- Membri del team (e team) diversi possono utilizzare gli stessi processi e procedure in modo coerente.
- I team procedono a scalare i processi tramite processi ripetibili.
- Processi e procedure standardizzati aiutano a mitigare l'impatto del trasferimento delle responsabilità del carico di lavoro tra i team.

Livello di rischio associato se questa best practice non fosse adottata: elevato

Guida all'implementazione

- Esistono proprietari identificati di processi e procedure, responsabili della loro definizione.
  - Identifica le attività operative eseguite a supporto dei carichi di lavoro. Documenta queste attività in un percorso individuabile.
  - Identifica in modo univoco la persona o il team responsabile della specifica di un'attività. Questo soggetto deve verificare la possibilità che questa possa essere correttamente eseguita dal componente di un team con opportune competenze, dotato di autorizzazioni, accesso e

strumenti adeguati. In caso di problemi nello svolgimento di tale attività, i membri del team che la eseguono sono responsabili della redazione di feedback dettagliati necessari per migliorarla.

- Acquisisci la proprietà dei metadati dell'elemento dell'attività tramite servizi come AWS Systems Manager, documenti e. AWS Lambda Acquisisci la responsabilità delle risorse utilizzando tag o gruppi di risorse, specificando proprietà e informazioni di contatto. Utilizzatelo AWS Organizations per creare politiche di etichettatura e acquisire informazioni sulla proprietà e sui contatti.

- Nel tempo, queste procedure si evolvono per essere eseguibili come codice, riducendo la necessità dell'intervento umano.
  - Ad esempio, prendete in considerazione AWS Lambda funzioni, CloudFormation modelli o documenti di automazione di AWS Systems Manager.
  - Esegui il controllo delle versioni nei repository appropriati.
  - Applica i tag adeguati alle risorse, in modo da agevolare l'identificazione di proprietari e documentazione.

Esempio del cliente

AnyCompany La vendita al dettaglio definisce la proprietà come il team o l'individuo che possiede i processi per un'applicazione o gruppi di applicazioni (che condividono pratiche e tecnologie architettoniche comuni). Inizialmente, il processo e le procedure sono documentati come step-by-step guide nel sistema di gestione dei documenti, individuabili tramite tag sul sistema Account AWS che ospita l'applicazione e su gruppi specifici di risorse all'interno dell'account. Fanno leva AWS Organizations per gestire i loro. Account AWS Nel tempo, questi processi vengono convertiti in codice e le risorse vengono definite utilizzando l'infrastruttura come codice (ad esempio i AWS Cloud Development Kit (AWS CDK) modelli CloudFormation o). I processi operativi diventano documenti di automazione in AWS Systems Manager o AWS Lambda funzioni, che possono essere avviati come attività pianificate, in risposta a eventi come AWS CloudWatch allarmi o AWS EventBridge eventi, o avviati da richieste all'interno di una piattaforma di gestione dei servizi IT (ITSM). Tutti i processi dispongono di tag per l'identificazione della proprietà. La documentazione per l'automazione e il processo viene mantenuta all'interno delle pagine wiki generate dal repository di codice per il processo.

Passaggi dell'implementazione

1. Documenta processi e procedure esistenti.
   a. Rivedili e conservali. up-to-date

   b. Identifica un proprietario per ciascun processo o procedura.
   c. Applica a ognuno il controllo delle versioni.
   d. Ove possibile, condividi processi e procedure tra carichi di lavoro e ambienti che condividono progetti architetturali.
2. Stabilisci meccanismi di feedback e miglioramento.
   a. Definisci policy relative alla frequenza di revisione dei processi.
   b. Definisci i processi per revisori e approvatori.
   c. Implementa i problemi o crea una coda di ticket per fornire e monitorare il feedback.
   d. Ove possibile, i processi e le procedure dovrebbero essere sottoposti all'approvazione preventiva e alla classificazione dei rischi da parte di un comitato per l'approvazione delle modifiche ()CAB.
3. Verifica che processi e procedure siano accessibili e individuabili da chi deve eseguirli.
   a. Utilizza i tag per indicare dove è possibile accedere a processi e procedure per il carico di lavoro.
   b. Utilizza messaggi di errore ed eventi significativi per indicare processi o procedure appropriati per risolvere un problema.
   c. Usa i wiki e la gestione dei documenti per rendere processi e procedure consultabili in modo coerente in tutta l'organizzazione.
4. Automatizza quando appropriato.
   a. Le automazioni dovrebbero essere sviluppate quando i servizi e le tecnologie forniscono un. API
   b. Fornisci indicazioni adeguate in merito ai processi. Sviluppa casi utente e requisiti per automatizzare i processi.
   c. Misura correttamente l'uso di processi e procedure e sfrutta i problemi come un'opportunità di miglioramento continuo.

Livello di impegno per il piano di implementazione: medio

Risorse

Best practice correlate:

- OPS02-BP01 Le risorse hanno identificato i proprietari
- OPS02-BP04 Esistono meccanismi per gestire le responsabilità e la proprietà
- OPS11-BP04 Esegui la gestione della conoscenza

Documenti correlati:

- AWS Whitepaper - Introduzione a on DevOps AWS
- AWS Whitepaper - Le migliori pratiche per l'etichettatura delle risorse AWS
- AWS Whitepaper - Organizzazione dell'ambiente utilizzando più account AWS
- Cloud AWS Blog sulle operazioni e le migrazioni - Sviluppa una pratica di automazione del cloud per l'eccellenza operativa: le migliori pratiche di AWS Managed Services
- Cloud AWS Blog sulle operazioni e le migrazioni - Implementazione di controlli di tagging automatizzati e centralizzati con e AWS ConfigAWS Organizations
- AWS Blog sulla sicurezza - Estendi i tuoi hook di pre-commit con AWS CloudFormation Guard
- AWS DevOps Blog - Integrazione AWS CloudFormation Guard nelle pipeline CI/CD

Workshop correlati:

- AWS Well-Architected Operational Excellence Workshop
- Workshop AWS - Tagging

Video correlati:

- Come automatizzare le operazioni IT su AWS
- AWS re:Invent 2020 - Automatizza qualsiasi cosa con Systems Manager AWS
- AWS re:Inforce 2022 - Automatizzazione della gestione e della conformità delle patch utilizzando (06) AWS NIS3
- AWS Support s You - Approfondimenti su AWS Systems Manager

Servizi correlati:

- AWS Systems Manager - Automazione
- AWS Service Management Connector

OPS02-BP03 Le attività operative hanno identificato i proprietari responsabili delle loro prestazioni

È utile sapere chi ha la responsabilità di eseguire attività specifiche su carichi di lavoro definiti e perché tale responsabilità esiste. Conoscere chi ha la responsabilità di eseguire le attività fornisce

indicazioni su chi eseguirà l'attività, chi convaliderà il risultato e chi fornirà feedback al proprietario dell'attività.

Risultato desiderato:

L'organizzazione definisce chiaramente le responsabilità per eseguire attività specifiche su carichi di lavoro stabiliti e rispondere agli eventi generati dai carichi di lavoro. L'organizzazione documenta la responsabilità dei processi e degli adempimenti e rende queste informazioni individuabili. Esamini e aggiorni le responsabilità in caso di cambiamenti organizzativi e i team monitorano e misurano le prestazioni delle attività di identificazione di difetti e inefficienze. Implementi i meccanismi di feedback per monitorare difetti e miglioramenti e supportare il miglioramento continuo.

Anti-pattern comuni:

- Mancata documentazione delle responsabilità.
- Presenza di script frammentati sulle workstation degli operatori isolate. Solo poche persone sanno come usarli o li chiamano informalmente conoscenze del team.
- Necessità di aggiornare un processo legacy, ma non si sa chi è il proprietario e l'autore originale non fa più parte dell'organizzazione.
- Mancata possibilità di individuare processi e script, quindi non sono immediatamente disponibili quando necessario, ad esempio, in risposta a un incidente.

Vantaggi dell'adozione di questa best practice:

- Sai chi è responsabile dell'esecuzione di un'attività, a chi notificare un'azione necessaria e chi esegue l'azione, convalida il risultato e fornisce il feedback al titolare dell'attività.
- Processi e procedure incentivano l'impegno nella gestione dei carichi di lavoro.
- I nuovi membri del team diventano efficienti in modo più rapido.
- Riduci il tempo necessario per mitigare gli incidenti.
- Team diversi utilizzano medesimi processi e procedure per eseguire le attività in modo coerente.
- I team procedono a scalare i processi tramite processi ripetibili.
- Processi e procedure standardizzati aiutano a mitigare l'impatto del trasferimento delle responsabilità del carico di lavoro tra i team.

Livello di rischio associato se questa best practice non fosse adottata: elevato

Guida all'implementazione

Per definire le responsabilità, inizia usando la documentazione esistente, ad esempio matrici di responsabilità, processi e procedure, ruoli e responsabilità, strumenti e automazione. Esamina la documentazione e organizza discussioni sulle responsabilità dei processi documentati. Collaborando con i team, identifica i disallineamenti tra le responsabilità e i processi documentati. Parla dei servizi offerti con i clienti interni dei team per identificare le divergenze nelle aspettative tra i team.

Analizza e risolvi le discrepanze. Identifica le opportunità di miglioramento e le attività richieste di frequente e con uso intensivo di risorse, in genere ottime candidate al miglioramento. Esamina best practice, modelli e linee guida prescrittive per semplificare e standardizzare i miglioramenti. Registra le opportunità di miglioramento e monitora i miglioramenti fino al completamento.

Nel tempo, queste procedure si evolvono per essere eseguibili come codice, riducendo la necessità dell'intervento umano. Ad esempio, le procedure possono essere avviate come AWS Lambda funzioni, AWS CloudFormation modelli o documenti di automazione. AWS Systems Manager Verifica che queste procedure siano sottoposte al controllo delle versioni nei repository appropriati e includano i corretti tag delle risorse in modo che i team possano identificare prontamente responsabili e documentazione. Documenta la responsabilità dello svolgimento delle attività, quindi monitora l'avvio e il funzionamento delle automazioni, nonché le prestazioni dei risultati desiderati.

Esempio del cliente

AnyCompany La vendita al dettaglio definisce la proprietà come il team o l'individuo che possiede i processi per un'applicazione o gruppi di applicazioni che condividono pratiche e tecnologie architettoniche comuni. Inizialmente, l'azienda documenta i processi e le procedure come step-by-step guide nel sistema di gestione dei documenti. Rendono le procedure individuabili utilizzando tag sull'ambiente Account AWS che ospita l'applicazione e su gruppi specifici di risorse all'interno dell'account, utilizzati AWS Organizations per gestirle Account AWS. Nel tempo, AnyCompany Retail converte questi processi in codice e definisce le risorse utilizzando l'infrastruttura come codice (tramite servizi CloudFormation o AWS Cloud Development Kit (AWS CDK) modelli). I processi operativi diventano documenti di automazione in AWS Systems Manager o nelle AWS Lambda funzioni, che possono essere avviati come attività pianificate in risposta a eventi come CloudWatch allarmi Amazon o EventBridge eventi Amazon o tramite richieste all'interno di una piattaforma di gestione dei servizi IT (ITSM). Tutti i processi dispongono dei tag per identificare il proprietario. I team gestiscono la documentazione per l'automazione e il processo nelle pagine wiki generate dal repository di codice per il processo.

Passaggi dell'implementazione

1. Documenta processi e procedure esistenti.
   a. Controlla e verifica che lo siano. up-to-date
   b. Verifica che ogni processo o procedura abbia un proprietario.
   c. Applica alle procedure il controllo delle versioni.
   d. Ove possibile, condividi processi e procedure tra carichi di lavoro e ambienti che condividono progetti architetturali.
2. Stabilisci meccanismi di feedback e miglioramento.
   a. Definisci policy relative alla frequenza di revisione dei processi.
   b. Definisci i processi per revisori e approvatori.
   c. Implementa i problemi o crea una coda di ticket per fornire e monitorare il feedback.
   d. Ove possibile, fornite la preapprovazione e la classificazione dei rischi per i processi e le procedure da parte di un comitato per l'approvazione delle modifiche (CAB).
3. Rendi i processi e le procedure accessibili e individuabili dagli utenti che devono eseguirli.
   a. Utilizza i tag per indicare dove è possibile accedere a processi e procedure per il carico di lavoro.
   b. Utilizza messaggi di errore ed eventi significativi per indicare il processo o la procedura appropriata per risolvere il problema.
   c. Usa i wiki o la gestione dei documenti per rendere i processi e le procedure consultabili in modo coerente in tutta l'organizzazione.
4. Automatizza quando è opportuno farlo.
   a. Laddove i servizi e le tecnologie forniscono e sviluppano automazioni. API
   b. Verifica che i processi siano ben compresi e sviluppa casi utente e requisiti per automatizzare i processi.
   c. Misura l'uso corretto di processi e procedure e sfrutta i problemi per supportare il miglioramento continuo.

Livello di impegno per il piano di implementazione: medio

Risorse

Best practice correlate:

- OPS02-BP01 Le risorse hanno identificato i proprietari
- OPS02-BP02 I processi e le procedure hanno identificato i proprietari
- OPS02-BP04 Esistono meccanismi per gestire le responsabilità e la proprietà
- OPS02-BP05 Esistono meccanismi per identificare la responsabilità e la titolarità
- OPS11-BP04 Esegui la gestione della conoscenza

Documenti correlati:

- AWS Whitepaper | Introduzione a on DevOps AWS
- AWS Whitepaper | Migliori pratiche per l'etichettatura delle risorse AWS
- AWS Whitepaper | Organizzazione dell'ambiente utilizzando più account AWS
- Cloud AWS Blog sulle operazioni e le migrazioni | Sviluppa una pratica di automazione del cloud per l'eccellenza operativa: le migliori pratiche di AWS Managed Services
- Workshop AWS - Tagging
- AWS Service Management Connector

Video correlati:

- AWS Knowledge Center Live | Risorse per l'etichettatura AWS
- AWS re:Invent 2020 | Automatizza qualsiasi cosa con Systems Manager AWS
- AWS re:Inforce 2022 | Automatizzazione della gestione e della conformità delle patch utilizzando (06) AWS NIS3
- AWS Support s You | Approfondimenti su AWS Systems Manager

Esempi correlati:

- AWS Well-Architected Operational Excellence Workshop

OPS02-BP04 Esistono meccanismi per gestire le responsabilità e la proprietà

Comprendi le responsabilità del tuo ruolo e il modo in cui contribuisci ai risultati aziendali in quanto questa conoscenza fornisce indicazioni sulle priorità delle tue attività e sul perché il tuo ruolo è importante. I membri del team possono quindi riconoscere le esigenze e rispondere in modo

appropriato. Quando i membri del team comprendono il proprio ruolo, possono stabilire la titolarità, identificare le opportunità di miglioramento e capire come influenzare o apportare le modifiche appropriate.

Occasionalmente, una responsabilità potrebbe non avere un titolare definito. In queste situazioni, progetta un meccanismo per risolvere la lacuna. Crea un percorso di escalation ben definito a qualcuno con l'autorità di assegnare la responsabilità o il piano per risolvere il problema.

Risultato desiderato: responsabilità definite in modo chiaro per i team all'interno dell'organizzazione, che comprendono il modo in cui sono correlate alle risorse, alle azioni da eseguire, ai processi e alle procedure. Queste responsabilità sono in linea con le responsabilità e gli obiettivi del team, nonché con le responsabilità degli altri team. Documenti i percorsi di escalation in modo coerente e individuabile e inserisci queste decisioni in artefatti di documentazione, come matrici di responsabilità, definizioni di team o pagine wiki.

Anti-pattern comuni:

- Le responsabilità del team sono ambigue o mal definite.
- Il team non allinea i ruoli alle responsabilità.
- Il team non allinea scopi e obiettivi alle responsabilità, rendendo difficile misurare il successo delle attività.
- Le responsabilità dei membri del team non sono in linea con il team e l'organizzazione in generale.
- Il team non mantiene le responsabilità up-to-date, il che le rende incoerenti con le attività svolte dal team.
- I percorsi di escalation per determinare le responsabilità non sono definiti o non sono chiari.
- I percorsi di escalation non hanno un unico responsabile del thread per garantire una risposta tempestiva.
- Ruoli, responsabilità e percorsi di escalation non sono individuabili e quindi non sono immediatamente disponibili quando richiesto, ad esempio in risposta a un incidente.

Vantaggi dell'adozione di questa best practice:

- Una volta compreso chi ha la responsabilità o la titolarità, puoi contattare il team o il membro del team appropriato per effettuare una richiesta o trasferire un'attività.
- Per ridurre il rischio di inattività e di esigenze non soddisfatte, identifichi una persona che ha l'autorità di assegnare responsabilità o titolarità.

- Quando si definisce chiaramente l'ambito di una responsabilità, i membri del team acquisiscono autonomia e titolarità.
- Le tue responsabilità forniscono indicazioni sulle decisioni che prendi, sulle azioni che intraprendi e sulle tue attività di distribuzione ai titolari appropriati.
- Ti sarà facile identificare le responsabilità abbandonate perché hai una chiara comprensione di ciò che non rientra nelle responsabilità del tuo team e quindi potrai effettuare l'escalation per chiedere chiarimenti.
- I team evitano confusione e tensione e possono gestire in modo più adeguato i carichi di lavoro e le risorse.

Livello di rischio associato se questa best practice non fosse adottata: elevato

Guida all'implementazione

Identifica i ruoli e le responsabilità dei membri del team e verifica che comprendano le aspettative del proprio ruolo. Rendi queste informazioni individuabili in modo che i membri della tua organizzazione possano identificare il team o la persona da contattare per esigenze specifiche. Man mano che le organizzazioni cercano di sfruttare le opportunità di migrazione e modernizzazione AWS, anche i ruoli e le responsabilità potrebbero cambiare. Rendi i team e i membri consapevoli delle loro responsabilità e offri la formazione appropriata per svolgere le attività durante questo cambiamento.

Determina il ruolo o il team che deve ricevere le escalation per identificare responsabilità e titolarità. Questo team può interagire con varie parti interessate per prendere le decisioni. Tuttavia, è proprietario della gestione del processo decisionale.

Fornisci ai membri della tua organizzazione meccanismi accessibili per scoprire e identificare titolarità e responsabilità. Questi meccanismi insegnano loro a chi rivolgersi per esigenze specifiche.

Esempio del cliente

AnyCompany La vendita al dettaglio ha recentemente completato una migrazione dei carichi di lavoro da un ambiente locale alla loro landing zone AWS con un approccio lift and shift. Ha eseguito una revisione delle operazioni per esaminare come vengono svolte le attività operative comuni e ha verificato che la matrice di responsabilità esistente rifletta le operazioni nel nuovo ambiente. Quando sono passati dall'ambiente locale a quello locale AWS, hanno ridotto le responsabilità dei team addetti all'infrastruttura relativa all'hardware e all'infrastruttura fisica. Questo passaggio ha anche rivelato nuove opportunità per evolvere il modello operativo dei carichi di lavoro.

Oltre ad aver identificato, risolto e documentato la maggior parte delle responsabilità, ha anche definito i percorsi di escalation per eventuali responsabilità mancanti o che potrebbero cambiare con l'evolversi delle procedure operative. Per esplorare nuove opportunità di standardizzare e migliorare l'efficienza dei carichi di lavoro, fornisci l'accesso a strumenti operativi come AWS Systems Manager e strumenti di sicurezza come AWS Security Hub Amazon. GuardDuty AnyCompanyRetail elabora una revisione delle responsabilità e della strategia sulla base dei miglioramenti che desidera apportare per primi. Man mano che l'azienda adotta nuovi modi di lavorare e modelli tecnologici, aggiorna la propria matrice di responsabilità di conseguenza.

Passaggi dell'implementazione

1. Inizia con la documentazione esistente. Alcuni documenti di origine tipici possono essere:
   a. Responsabilità o matrici responsabili, responsabili, consultate e informate () RACI
   b. Definizioni dei team o pagine wiki.
   c. Definizioni e offerte di servizi.
   d. Ruolo o descrizione delle mansioni lavorative.
2. Esamina la documentazione e organizza discussioni sulle responsabilità documentate:
   a. Collaborando con i team identifica i disallineamenti tra le responsabilità documentate e quelle normalmente assunte dai team.
   b. Esamina i potenziali servizi offerti dai clienti interni per identificare le lacune nelle aspettative tra i team.
3. Analizza e risolvi le discrepanze.
4. Identifica le opportunità di miglioramento.
   a. Identifica le richieste più frequenti e con uso intensivo di risorse, che in genere sono ottime candidate al miglioramento.
   b. Esamina le best practice, i modelli e le linee guida prescrittive per semplificare e standardizzare i miglioramenti.
   c. Registra le opportunità di miglioramento e monitorale fino al completamento.
5. Se nessuno nel team è responsabile della gestione e del monitoraggio dell'assegnazione delle responsabilità, identifica qualcuno che assuma tale responsabilità.
6. Definisci un processo per consentire ai team di richiedere chiarimenti sulla responsabilità.
   a. Esamina il processo e verifica che sia chiaro e semplice da usare.
   b. Assicurati che qualcuno sia proprietario e segua le escalation fino al completamento.
   c. Stabilisci le metriche operative per misurare l'efficacia.

   d. Crea un meccanismo di feedback per verificare che i team possano evidenziare le opportunità di miglioramento.
   e. Implementa un meccanismo di revisione periodica.
7. Rendi i documenti disponibili in una posizione individuabile e accessibile.
   a. I wiki o il portale di documentazione sono le posizioni normalmente scelte.

Livello di impegno per il piano di implementazione: medio

Risorse

Best practice correlate:

- OPS01-BP06 Valuta i compromessi
- OPS03-BP02 I membri del team hanno il potere di agire quando i risultati sono a rischio
- OPS03-BP03 L'escalation è incoraggiata
- OPS03-BP07 I team addetti alle risorse sono appropriati
- OPS09-BP01 Misura gli obiettivi operativi e con le metriche KPIs
- OPS09-BP03 Rivedi le metriche operative e dai priorità al miglioramento
- OPS11-BP01 Adottate un processo per il miglioramento continuo

Documenti correlati:

- AWS Whitepaper - Introduzione a on DevOps AWS
- AWS Whitepaper - Quadro di adozione: prospettiva operativa Cloud AWS
- Eccellenza operativa del Framework AWS Well-Architected: topologie del modello operativo a livello di carico di lavoro
- AWS Prescriptive Guidance - Building your Cloud Operating Model
- AWS Guida prescrittiva: crea una RASCI matrice RACI or per un modello operativo cloud
- Cloud AWS Blog sulle operazioni e le migrazioni - Offrire valore aziendale con i team della piattaforma cloud
- Cloud AWS Blog sulle operazioni e le migrazioni - Perché un modello operativo cloud?
- AWS DevOps Blog - In che modo le organizzazioni si stanno modernizzando per le operazioni sul cloud

Video correlati:

- AWS Summit Online - Cloud Operating Models for Accelerated Transformation
- AWS re:Invent 2023 - Future-proofing cloud security: A new operating model

OPS02-BP05 Esistono meccanismi per richiedere aggiunte, modifiche ed eccezioni

È possibile effettuare richieste ai titolari di processi, procedure e risorse. Tra le richieste figurano aggiunte, modifiche ed eccezioni. Tali richieste passano attraverso un processo di gestione delle modifiche Prendi decisioni informate per approvare le richieste quando vengono ritenute fattibili e appropriate dopo una valutazione dei vantaggi e dei rischi.

Risultato desiderato:

- Puoi effettuare richieste per modificare processi, procedure e risorse sulla base della titolarità assegnata.
- Le modifiche vengono eseguite in modo deliberato, valutando benefici e rischi.

Anti-pattern comuni:

- Devi aggiornare il modo di implementare la tua applicazione, ma non esiste un metodo per richiedere una modifica al processo di implementazione al team operativo.
- Il piano di ripristino di emergenza deve essere aggiornato, ma non è stato identificato il proprietario a cui richiedere le modifiche.

Vantaggi dell'adozione di questa best practice:

- Processi, procedure e risorse possono evolvere mentre cambiano i requisiti.
- I titolati possono prendere decisioni mirate su quando effettuare le modifiche.
- Le modifiche vengono eseguite deliberatamente.

Livello di rischio associato se questa best practice non fosse adottata: medio

Guida all'implementazione

Per implementare questa best practice devi essere in grado di richiedere modifiche a processi, procedure e risorse. Il processo di gestione delle modifiche può essere semplice. Documenta il processo di gestione delle modifiche.

Esempio del cliente

AnyCompany Retail utilizza una matrice di assegnazione delle responsabilità (RACI) per identificare i responsabili delle modifiche relative a processi, procedure e risorse. L'azienda dispone di un processo documentato di gestione delle modifiche, semplice e facile da seguire. Utilizzando la RACI matrice e il processo, chiunque può inviare richieste di modifica.

Passaggi dell'implementazione

1. Identifica i processi, le procedure e le risorse per il tuo carico di lavoro e i proprietari di ciascun elemento. Documentali nel tuo sistema di gestione delle conoscenze.
   a. In caso di mancata implementazione, inizia da OPS02-BP01 Le risorse hanno identificato i proprietari, OPS02-BP02 I processi e le procedure hanno identificato i proprietari o OPS02-BP03 Le attività operative hanno identificato i proprietari responsabili delle loro prestazioni.
2. Collabora con le parti interessate all'interno della tua azienda per sviluppare un processo di gestione delle modifiche. Il processo deve includere aggiunte, modifiche ed eccezioni per risorse, processi e procedure.
   a. Puoi utilizzare AWS Systems Manager Change Manager come piattaforma di gestione delle modifiche per le risorse del carico di lavoro.
3. Documenta il processo di gestione delle modifiche nel tuo sistema di gestione delle conoscenze.

Livello di impegno per il piano di implementazione: medio Sviluppare un processo di gestione delle modifiche significa garantire un allineamento con più parti interessate all'interno dell'organizzazione.

Risorse

Best practice correlate:

- OPS02-BP01 Le risorse hanno identificato i proprietari: le risorse richiedono proprietari identificati prima di creare un processo di gestione delle modifiche.
- OPS02-BP02 I processi e le procedure hanno identificato i proprietari: i processi richiedono proprietari identificati prima di creare un processo di gestione delle modifiche.

- OPS02-BP03 Le attività operative hanno identificato i proprietari responsabili delle loro prestazioni: le attività di operazioni richiedono proprietari identificati prima di creare un processo di gestione delle modifiche.

Documenti correlati:

- AWS Guida prescrittiva - Playbook di base per migrazioni di AWS grandi dimensioni: creazione di matrici RACI
- Whitepaper sulla gestione delle modifiche nel cloud

Servizi correlati:

- AWS Systems Manager Change Manager

OPS02-BP06 Le responsabilità tra i team sono predefinite o negoziate

Predisponi accordi definiti o concordati tra i team che descrivono come funzionano e si supportano reciprocamente (ad esempio, tempi di risposta, obiettivi o contratti relativi al livello di servizio). I canali di comunicazione tra team sono documentati. Comprendere l'impatto del lavoro dei team sui risultati aziendali e sui risultati di altri team e organizzazioni fornisce indicazioni in merito alla priorità dei loro compiti e consente loro di rispondere in modo appropriato.

Quando la responsabilità e la proprietà sono indefinite o sconosciute, rischi di non affrontare le attività necessarie in modo tempestivo e di impiegare sforzi ridondanti e potenzialmente conflittuali per rispondere a tali esigenze.

Risultato desiderato:

- Il lavoro tra team o gli accordi di assistenza vengono concordati e documentati.
- I team che supportano o lavorano con altri hanno definito i canali di comunicazione e le aspettative in termini di risposte.

Anti-pattern comuni:

- In produzione si verifica un problema e due team separati iniziano a cercare la soluzione senza confrontarsi. Il loro impegno separato prolunga l'interruzione.

- Il team operativo ha bisogno di assistenza dal team di sviluppo, ma non c'è un accordo sui tempi di risposta. La richiesta si blocca nel backlog.

Vantaggi dell'adozione di questa best practice:

- I team sanno come interagire e supportarsi a vicenda.
- Le aspettative relative ai tempi di risposta sono note.
- I canali di comunicazione sono definiti in modo chiaro.

Livello di rischio associato se questa best practice non fosse adottata: basso

Guida all'implementazione

Se si implementa questa best practice non ci saranno dubbi sulla collaborazione tra team. Gli accordi formali codificano il modo di collaborare o di assistersi a vicenda dei team. I canali di comunicazione tra team sono documentati.

Esempio del cliente

AnyCompany Il SRE team di Retail ha stipulato un accordo sul livello di servizio con il team di sviluppo. Ogni volta che il team di sviluppo effettua una richiesta nel sistema di ticketing, riceve una risposta entro 15 minuti. In caso di interruzione del sito, il SRE team si occupa delle indagini con il supporto del team di sviluppo.

Passaggi dell'implementazione

1. Collaborando con le parti interessate all'interno dell'organizzazione, sviluppa accordi tra team basati su processi e procedure.
   a. Se i due team condividono un processo o una procedura, crea un runbook sulle modalità di collaborazione dei team.
   b. Se ci sono dipendenze tra i team, accetta una risposta alle SLA richieste.
2. Inserisci le responsabilità nel tuo sistema di gestione delle conoscenze.

Livello di impegno per il piano di implementazione: medio Se non esistono accordi tra i team, può essere impegnativo raggiungere un accordo con le parti interessate all'interno dell'organizzazione.

Risorse

Best practice correlate:

- OPS02-BP02 I processi e le procedure hanno identificato i proprietari: la proprietà del processo deve essere identificata prima di stabilire accordi tra i team.
- OPS02-BP03 Le attività operative hanno identificato i proprietari responsabili delle loro prestazioni: la proprietà delle operazioni deve essere identificata prima di stabilire accordi tra i team.

Documenti correlati:

- AWS Executive Insights - Potenziare l'innovazione con il team di Two-Pizza
- Introduzione a DevOps on AWS - Two-Pizza Teams

OPS3. In che modo la cultura aziendale supporta i risultati aziendali?

Fornisci supporto ai membri del team in modo che possano essere più efficaci nell'azione e nel supporto dei risultati aziendali.

Best practice

- OPS03-BP01 Fornire una sponsorizzazione esecutiva
- OPS03-BP02 I membri del team hanno il potere di agire quando i risultati sono a rischio
- OPS03-BP03 L'escalation è incoraggiata
- OPS03-BP04 Le comunicazioni sono tempestive, chiare e utilizzabili
- OPS03-BP05 La sperimentazione è incoraggiata
- OPS03-BP06 I membri del team sono incoraggiati a mantenere e accrescere le proprie competenze
- OPS03-BP07 Team di risorse appropriati

OPS03-BP01 Fornire una sponsorizzazione esecutiva

Ai massimi livelli, gli alti dirigenti fungono da sponsor esecutivo per definire chiaramente le aspettative e la direzione dei risultati dell'organizzazione, compresa la valutazione del successo. Lo sponsor sostiene e promuove l'adozione delle best practice e l'evoluzione dell'organizzazione.

Risultato desiderato: definizione di linee chiare in termini di leadership e responsabilità per i risultati desiderati da parte delle organizzazioni impegnate nell'adottare, trasformare e ottimizzare le proprie operazioni cloud. L'organizzazione comprende ogni capacità richiesta per raggiungere un nuovo risultato e assegna la proprietà ai team funzionali per lo sviluppo. La leadership stabilisce attivamente questa direzione, assegna la proprietà, si assume la responsabilità e definisce il lavoro. Di conseguenza, le persone in tutta l'organizzazione possono mobilitarsi, sentirsi ispirate e lavorare attivamente per raggiungere gli obiettivi desiderati.

Anti-pattern comuni:

- I proprietari dei carichi di lavoro sono tenuti a migrare i carichi di lavoro su AWS senza uno sponsor e un piano chiari per le operazioni cloud. I team pertanto non collaborano in modo consapevole per migliorare e consolidare le proprie capacità operative. La mancanza di standard operativi sulle best practice mette in difficoltà i team, ad esempio il lavoro degli operatori, le chiamate e il debito tecnico, limitando l'innovazione.
- È stato fissato un nuovo obiettivo a livello di organizzazione per adottare una tecnologia emergente senza fornire sponsor e strategia di leadership. I team interpretano gli obiettivi in modo diverso, il che crea confusione su dove concentrare gli impegni, sul perché sono importanti e su come misurare l'impatto. Di conseguenza, l'organizzazione perde slancio nell'adozione della tecnologia.

Vantaggi dell'adozione di questa best practice: se lo sponsor esecutivo comunica e condivide in modo chiaro visione, direzione e obiettivi, i membri del team conoscono le aspettative riposte su di loro. Quando i leader sono coinvolti attivamente, le persone e i team iniziano a concentrare attivamente gli impegni nella stessa direzione per raggiungere gli obiettivi definiti. L'organizzazione di conseguenza massimizza la capacità di successo. Quando si valuta il successo, è possibile identificare meglio gli ostacoli al suo conseguimento in modo da affrontarli attraverso l'intervento dello sponsor esecutivo.

Livello di rischio associato se questa best practice non fosse adottata: elevato

Guida all'implementazione

- In ogni fase del percorso verso il cloud (migrazione, adozione oppure ottimizzazione), il successo richiede un coinvolgimento attivo ai massimi livelli della leadership con uno sponsor esecutivo designato. Lo sponsor esecutivo allinea la mentalità, le competenze e le modalità di lavoro del team alla strategia definita.
  - Spiega il perché: chiarisci e illustra il ragionamento alla base di visione e strategia.

- Definisci le aspettative: definisci e pubblica gli obiettivi per le tue organizzazioni, incluso il modo in cui verranno misurati.
- Tieni traccia del conseguimento degli obiettivi: misura con regolarità il conseguimento incrementale degli obiettivi (non solo il completamento delle attività). Condividi i risultati in modo da poter intraprendere le azioni appropriate se si evidenziano dei rischi.
- Fornisci le risorse necessarie per raggiungere gli obiettivi: favorisci la collaborazione tra persone e team al fine di sviluppare le soluzioni giuste che garantiscano i risultati definiti. Ciò riduce o elimina gli attriti organizzativi.
- Sostieni i team: mantieni un coinvolgimento attivo con i tuoi team in modo da comprenderne le prestazioni e l'eventuale presenza di fattori di influenza esterni. Individua gli ostacoli che impediscono i progressi dei team. Agisci per conto dei tuoi team per superare gli ostacoli e rimuovere gli oneri superflui. Quando i team sono influenzati da fattori esterni, rivaluta gli obiettivi e modifica i target in base alle esigenze.
- Promuovi l'adozione delle best practice: riconosci le best practice che offrono vantaggi quantificabili e identifica creatori e destinatari. Incoraggia ulteriormente l'adozione per amplificare i vantaggi ottenuti.
- Incoraggia l'evoluzione dei team: crea una cultura di miglioramento continuo e impara in modo proattivo da progressi e insuccessi. Incoraggia la crescita e lo sviluppo sia personale sia organizzativo. Usa dati e aneddoti per migliorare la visione e la strategia.

Esempio del cliente

AnyCompany La vendita al dettaglio sta attraversando un processo di trasformazione aziendale attraverso la rapida reinvenzione delle esperienze dei clienti, il miglioramento della produttività e l'accelerazione della crescita attraverso l'IA generativa.

Passaggi dell'implementazione

1. Stabilisci una leadership a thread singolo e assegna uno sponsor esecutivo principale per guidare e gestire la trasformazione.
2. Definisci chiaramente i risultati aziendali della trasformazione e assegna proprietà e responsabilità. Fornisci allo sponsor esecutivo principale l'autorità di guidare e prendere decisioni critiche.
3. Verifica che la strategia di trasformazione sia stata definita molto chiaramente e ampiamente comunicata dallo sponsor esecutivo a tutti i livelli dell'organizzazione.
   a. Definisci chiaramente gli obiettivi aziendali per le iniziative IT e cloud.

   b. Documenta le principali metriche aziendali per promuovere la trasformazione dell'IT e del cloud.
   c. Comunica la visione in modo coerente a tutti i team e alle persone responsabili di parti della strategia.
4. Sviluppa matrici di pianificazione della comunicazione che specifichino quale messaggio deve essere recapitato a leader, manager e singoli collaboratori specifici. Specifica la persona o il team che deve recapitare questo messaggio.
   a. Rispetta i piani di comunicazione in modo coerente e affidabile.
   b. Stabilisci e gestisci le aspettative attraverso eventi di persona su base regolare.
   c. Accetta il feedback sull'efficacia delle comunicazioni, quindi modifica le comunicazioni e pianifica di conseguenza.
   d. Pianifica gli eventi di comunicazione per comprendere in modo proattivo le sfide dei team e stabilire un ciclo di feedback coerente che consenta di correggere la direzione laddove necessario.
5. Coinvolgi in modo attivo ogni iniziativa dal punto di vista della leadership per verificare che tutti i team interessati comprendano i risultati di cui sono responsabili.
6. In ogni riunione sullo stato, gli sponsor esecutivi devono individuare gli ostacoli, esaminare metriche, aneddoti o feedback dei team nonché misurare i progressi verso il raggiungimento degli obiettivi.

Livello di impegno per il piano di implementazione: medio

Risorse

Best practice correlate:

- OPS03-BP04 Le comunicazioni sono tempestive, chiare e utilizzabili
- OP11-BP01 Adottate un processo per il miglioramento continuo
- OPS11-BP07 Esegui revisioni delle metriche operative

Documenti correlati:

- Untangling Your Organisational Hairball: Highly Aligned
- The Living Transformation: Pragmatically approaching changes
- Becoming a Future-Ready Enterprise

- 7 insidie da evitare quando si costruisce un CCOE
- Navigating the Cloud: Key Performance Indicators for Success

Video correlati:

- AWS re:Invent 2023: Una guida per i leader all'intelligenza artificiale generativa: usare la storia per plasmare il futuro (04) SEG2

Esempi correlati:

- Prosci: Primary Sponsor's Role & Importance

OPS03-BP02 I membri del team hanno il potere di agire quando i risultati sono a rischio

Il comportamento culturale della responsabilità instillato dalla leadership fa sì che ogni dipendente si senta autorizzato ad agire per conto dell'intera azienda, al di là del proprio ambito definito da ruolo e responsabilità. I dipendenti possono intervenire per identificare in modo proattivo i rischi man mano che emergono e intraprendere le azioni appropriate. Tale cultura consente ai dipendenti di prendere decisioni di alto valore in quanto consapevoli della situazione.

Ad esempio, Amazon utilizza i principi di leadership come linee guida per agevolare comportamenti migliori dei dipendenti nelle situazioni, la risoluzione dei problemi, l'affrontare i conflitti e l'agire.

Risultato desiderato: una nuova cultura stabilita dalla leadership che consente a persone e di prendere decisioni critiche, anche ai livelli inferiori dell'organizzazione (a condizione che le decisioni a lungo termine siano definite con autorizzazioni e meccanismi di sicurezza sottoponibili ad audit). L'errore non è una mancanza, i team imparano in modo iterativo a migliorare il processo decisionale e le risposte per affrontare situazioni simili in futuro. Se le azioni già intraprese portano a un miglioramento che può avvantaggiare altri team, occorre condividere in modo proattivo le conoscenze derivanti da tali azioni. La leadership misura i miglioramenti operativi e incentiva le persone e l'organizzazione all'adozione di tali modelli.

Anti-pattern comuni:

- Nell'organizzazione non esistono linee guida o meccanismi chiari su cosa fare quando viene identificato un rischio. Ad esempio, quando un dipendente nota un attacco di phishing, non lo segnala al team di sicurezza, con il risultato che gran parte dell'organizzazione è vittima dell'attacco, causando una violazione dei dati.

- I clienti si lamentano dell'indisponibilità del servizio, che deriva principalmente da implementazioni non riuscite. Il tuo SRE team è responsabile dello strumento di implementazione e il rollback automatizzato delle implementazioni rientra nella loro tabella di marcia a lungo termine. In un recente rollout dell'applicazione, uno degli ingegneri ha fornito una soluzione per automatizzare il ripristino dell'applicazione a una versione precedente. Sebbene la loro soluzione possa diventare un modello per i SRE team, altri team non la adottano, in quanto non esiste un processo che consenta di tenere traccia di tali miglioramenti. L'organizzazione continua a essere afflitta da implementazioni non corrette che hanno un impatto sui clienti e provocano ulteriore insoddisfazione.
- Per garantire la conformità, il tuo team Infosec supervisiona un processo consolidato che prevede la rotazione regolare delle SSH chiavi condivise per conto degli operatori che si connettono alle loro istanze Amazon Linux. EC2 I team di infosec impiegano diversi giorni per completare la rotazione delle chiavi e la connessione alle istanze viene bloccata. Nessuno, interno o esterno a infosec, suggerisce di utilizzare altre opzioni per ottenere lo stesso risultato. AWS

Vantaggi dell'adozione di questa best practice: la decentralizzazione dell'autorità per l'adozione delle decisioni e la concessione ai team della possibilità di adottare decisioni chiave consente di affrontare i problemi più rapidamente, con percentuali di successo crescenti. Inoltre, i team iniziano a percepire un senso di appartenenza e gli errori sono accettabili. La sperimentazione diventa un pilastro culturale. Manager e direttori non si sentono controllati in ogni aspetto del loro lavoro.

Livello di rischio associato se questa best practice non fosse adottata: medio

Guida all'implementazione

1. Sviluppa una cultura che preveda il verificarsi errori.
2. Definisci chiaramente proprietà e responsabilità per le varie aree funzionali all'interno dell'organizzazione.
3. Comunica la proprietà e la responsabilità a tutti in modo che le persone sappiano chi può facilitare le decisioni decentralizzate.
4. Stabilisci le decisioni definitive e reversibili per permettere alle persone di sapere quando è necessario eseguire l'escalation a livelli più alti di leadership.
5. Crea la consapevolezza organizzativa secondo cui tutti i dipendenti hanno la capacità di agire a vari livelli quando i risultati sono a rischio. Fornisci ai membri del team la documentazione sulla governance, i livelli di autorizzazione, gli strumenti e le opportunità per mettere in pratica le competenze necessarie e intervenire in modo efficace.

6. Offri ai membri del team l'opportunità di mettere in pratica le competenze necessarie per rispondere a varie decisioni. Una volta definiti i livelli decisionali, organizza delle giornate di gioco per verificare che tutti i singoli collaboratori comprendano e possano usare il processo.
   a. Fornisci ambienti sicuri alternativi in cui testare i processi e sottoporre i membri del team alla dovuta formazione.
   b. Riconosci e crea la consapevolezza secondo cui i membri del team hanno l'autorità di agire quando il risultato ha un livello di rischio prestabilito.
   c. Definisci l'autorità dei membri del team per intervenire assegnando le autorizzazioni e l'accesso ai carichi di lavoro e ai componenti supportati.
7. Offri ai team la possibilità di condividere le proprie conoscenze (successi e fallimenti operativi).
8. Consenti ai team di sfidare lo status quo e fornisci i meccanismi per monitorare e misurare i miglioramenti, nonché il loro impatto sull'organizzazione.

Livello di impegno per il piano di implementazione: medio

Risorse

Best practice correlate:

- OPS01-BP06 Valuta i compromessi gestendo vantaggi e rischi
- OPS02-BP05 Esistono meccanismi per identificare la responsabilità e la proprietà

Documenti correlati:

- Post sul blog AWS | The agile enterprise
- Post sul blog AWS | Measuring success: A paradox and a plan
- Post sul blog AWS | Letting go: Enabling autonomy in teams
- Centralize or Decentralize?

Video correlati:

- re:Invent 2023 | Come non sabotare la trasformazione (01) SEG2
- re:Invent 2021 | Amazon Builders' Library: Operational Excellence at Amazon
- Centralization vs. Decentralization

Esempi correlati:

- Using architectural decision records to streamline technical decision-making for a software development project

OPS03-BP03 L'escalation è incoraggiata

I membri del team sono incoraggiati dalla leadership a segnalare problemi e preoccupazioni ai responsabili delle decisioni e alle parti interessate di alto livello se ritengono che i risultati desiderati siano a rischio e gli standard previsti non siano rispettati. Questa è una funzionalità della cultura dell'organizzazione ed è implementata a tutti i livelli. L'escalation deve essere eseguita in anticipo e di frequente, in modo da identificare i rischi e limitarli prima che provochino incidenti. La leadership non rimprovera le persone per aver effettuato l'escalation di un problema.

Risultato desiderato: possibilità per le persone in tutta di eseguire l'escalation dei problemi ai loro livelli di leadership immediati e superiori. La leadership ha stabilito deliberatamente e consapevolmente l'aspettativa che i propri team si sentano tranquilli nell'eseguire l'escalation di qualsiasi problema. Esiste un meccanismo per eseguire l'escalation dei problemi a ogni livello dell'organizzazione. Quando un dipendente esegue l'escalation al proprio manager, insieme decidono il livello di impatto e se il problema debba essere ulteriormente scalato. Per iniziare l'escalation, i dipendenti sono tenuti a includere un piano di lavoro consigliato per risolvere il problema. Se la direzione non interviene tempestivamente, i dipendenti sono incoraggiati a inoltrare i problemi al massimo livello di leadership se ritengono fermamente che i rischi per l'organizzazione giustifichino l'escalation.

Anti-pattern comuni:

- I dirigenti non pongono domande approfondite durante la riunione sullo stato del programma di trasformazione del cloud per scoprire dove si verificano problemi e ostacoli. Solo le buone notizie vengono presentate nello stato. The CIO ha chiarito che le piace solo sentire buone notizie, poiché qualsiasi sfida sollevata fa CEO pensare che il programma stia fallendo.
- Sei un ingegnere delle operazioni cloud e noti che il nuovo sistema di gestione delle conoscenze non è ampiamente adottato dai team applicativi. L'azienda ha investito un anno di tempo e diversi milioni di dollari per implementare questo nuovo sistema di gestione delle conoscenze, ma le persone continuano a creare i propri runbook localmente e a condividerli su una condivisione cloud aziendale, rendendo difficile l'individuazione delle conoscenze pertinenti ai carichi di lavoro supportati. Cerchi di portare questo aspetto all'attenzione della dirigenza perché l'uso coerente del sistema può migliorare l'efficienza operativa. Quando lo comunichi alla direttrice a capo

dell'implementazione del sistema di gestione delle conoscenze, ti rimprovera perché tale aspetto mette in discussione l'investimento.

- Il team di infosec responsabile del rafforzamento delle risorse di elaborazione ha deciso di mettere in atto un processo che richiede l'esecuzione delle scansioni necessarie per garantire che le EC2 istanze siano completamente protette prima che il team di elaborazione rilasci la risorsa per l'uso. Ciò ha comportato un ritardo di un'ulteriore settimana per l'implementazione delle risorse, il che interrompe il loro periodo di tempo. SLA Il team di calcolo non desidera inoltrare la questione al vicepresidente tramite cloud perché ciò mette in cattiva luce il vicepresidente della sicurezza delle informazioni.

Vantaggi dell'adozione di questa best practice:

I problemi complessi o critici vengono risolti prima che abbiano impatto sull'azienda. Si perde meno tempo. I rischi sono ridotti al minimo. I team diventano più proattivi e concentrati sui risultati della risoluzione dei problemi.

Livello di rischio associato se questa best practice non fosse adottata: elevato

Guida all'implementazione

La volontà e la capacità di crescere liberamente a tutti i livelli dell'organizzazione sono la base organizzativa e culturale da sviluppare consapevolmente attraverso una formazione appropriata, le comunicazioni della leadership, la definizione delle aspettative e l'implementazione di meccanismi a tutti i livelli dell'organizzazione.

Passaggi dell'implementazione

1. Definisci policy, standard e aspettative per l'organizzazione.
   a. Garantisci un'ampia adozione e comprensione delle policy, delle aspettative e degli standard.
2. Incoraggia, forma e responsabilizza i lavoratori a eseguire un'escalation anticipata e frequente quando gli standard non vengono rispettati.
3. Riconosci a livello organizzativo che l'escalation anticipata e frequente è la best practice. Accetti che le escalation possono rivelarsi infondate e che è meglio avere l'opportunità di prevenire un incidente piuttosto che privarsi di quell'opportunità senza escalation.
   a. Predisponi un meccanismo di escalation (come un sistema Andon cord).
   b. È opportuno disporre di procedure documentate che definiscano quando e come deve verificarsi l'escalation.

   c. Definisci la serie di persone in ordine di autorità cui è consentito intraprendere o approvare azioni, nonché le informazioni di contatto di ciascuna parte interessata.
4. Un'escalation deve continuare fino a quando il membro del team non è convinto che il rischio sia stato mitigato attraverso le azioni guidate dalla leadership.
   a. Le escalation devono includere:
      i. la descrizione della situazione e la natura del rischio;
      ii. le criticità della situazione;
      iii. chi o cosa è interessato;
      iv. il livello dell'impatto;
      v. l'urgenza in caso di impatto;
      vi. i rimedi suggeriti e i piani di mitigazione.
   b. Proteggi i dipendenti coinvolti nell'escalation. È necessario predisporre una policy che protegga i membri del team da eventuali ritorsioni se si trovano a dover scavalcare una parte interessata o un responsabile delle decisioni non reattivo. Metti in atto dei meccanismi per identificare se ciò si verifica e rispondere in modo appropriato.
5. Incoraggia la cultura del miglioramento continuo e dei cicli di feedback in tutto ciò che l'organizzazione produce. I cicli di feedback fungono da piccole escalation per le persone responsabili e identificano le opportunità di miglioramento, anche quando l'escalation non è necessaria. La cultura del miglioramento continuo obbliga tutti a essere più proattivi.
6. La leadership deve periodicamente ribadire le policy, gli standard, i meccanismi e il desiderio di un'escalation aperta e di cicli di feedback continui senza penalità.

Livello di impegno per il piano di implementazione: medio

Risorse

Best practice correlate:

- OPS02-BP05 Esistono meccanismi per richiedere aggiunte, modifiche ed eccezioni

Documenti correlati:

- How do you foster a culture of continuous improvement and learning from Andon and escalation systems?
- The Andon Cord (IT Revolution)

- AWS DevOps Linee guida | Stabilisci percorsi di escalation chiari e incoraggia un disaccordo costruttivo

Video correlati:

- Jeff Bezos on how to make decisions (& increase velocity)
- Toyota Product System: Stopping Production, a Button, and an Andon Electric Board
- Andon Cord nel settore della produzione LEAN

Esempi correlati:

- Working with escalation plans in Incident Manager

OPS03-BP04 Le comunicazioni sono tempestive, chiare e utilizzabili

La leadership è responsabile della creazione di comunicazioni forti ed efficaci, soprattutto quando l'organizzazione adotta nuove strategie, tecnologie o modalità di lavoro. I leader devono stabilire le aspettative affinché tutto il personale lavori per raggiungere gli obiettivi aziendali. Elabora meccanismi di comunicazione che creino e mantengano la consapevolezza tra i team responsabili della gestione dei piani finanziati e sponsorizzati dalla leadership. Utilizza la diversità interorganizzativa e ascolta con attenzione i vari punti di vista. Usa questa prospettiva per incrementare l'innovazione, mettere in discussione le tue ipotesi e ridurre il rischio di bias confermativi. Favorisci l'inclusione, la diversità e l'accessibilità all'interno dei team per ottenere prospettive vantaggiose.

Risultato desiderato: elaborazione di strategie di comunicazione da parte della tua organizzazione per gestire l'impatto del cambiamento sull'organizzazione. I team sono informati e motivati a continuare a lavorare insieme anziché l'uno contro l'altro. Le persone comprendono quanto sia importante il proprio ruolo per raggiungere gli obiettivi stabiliti. L'e-mail è solo un meccanismo passivo per le comunicazioni e viene utilizzato di conseguenza. La direzione trascorre tempo con i singoli collaboratori per aiutarli a comprendere le proprie responsabilità, le attività da completare e in che modo il loro lavoro contribuisce alla missione generale. Quando necessario, i leader coinvolgono direttamente le persone in un ambiente più piccolo per trasmettere il messaggio e verificare che venga recepito in modo efficace. Come risultato di buone strategie di comunicazione, l'organizzazione si comporta in misura pari o superiore alle aspettative della leadership. La leadership incoraggia e desidera esaminare opinioni diverse all'interno dell'organizzazione e tra i team.

Anti-pattern comuni:

- L'organizzazione ha un piano quinquennale per migrare tutti i carichi di lavoro su AWS. Il business case per il cloud include la modernizzazione del 25% di tutti i carichi di lavoro per utilizzare la tecnologia serverless. CIOComunica questa strategia ai referenti diretti e si aspetta che ogni leader trasmetta questa presentazione a cascata a manager, direttori e singoli collaboratori senza alcuna comunicazione di persona. Fa un CIO passo indietro e si aspetta che la sua organizzazione attui la nuova strategia.
- La leadership non fornisce né utilizza un meccanismo di feedback e aumenta il divario nelle aspettative, causando lo stallo dei progetti.
- Ti viene chiesto di apportare una modifica ai gruppi di sicurezza, ma non ricevi i dettagli sulle stesse, sull'impatto della modifica su tutti i carichi di lavoro e sulla data della modifica. Il manager inoltra un'e-mail dal vicepresidente di InfoSec e aggiunge il messaggio «Fai in modo che accada».
- Sono state apportate modifiche alla strategia di migrazione che riducono la percentuale di modernizzazione pianificata dal 25% al 10%. La riduzione ha effetti a valle sull'organizzazione delle operazioni. Questo cambiamento strategico non è stato comunicato e quindi non è disponibile la capacità qualificata sufficiente per supportare un numero maggiore di carichi di lavoro in lift and shift in AWS.

Vantaggi dell'adozione di questa best practice:

- L'organizzazione è ben informata sulle strategie nuove o modificate e agisce di conseguenza con una forte motivazione alla collaborazione per raggiungere gli obiettivi e le metriche generali stabiliti dalla leadership.
- Esistono meccanismi utilizzati per fornire tempestivamente notifiche ai membri del team in merito a rischi noti ed eventi pianificati.
- Le nuove modalità di lavoro, compresi i cambiamenti relativi a personale o organizzazione, processi o tecnologia, insieme alle competenze richieste, vengono adottate in modo più efficace dall'organizzazione che quindi realizza i vantaggi aziendali più rapidamente.
- I membri del team hanno il contesto necessario per ricevere le comunicazioni e possono essere più efficaci nel loro lavoro.

Livello di rischio associato se questa best practice non fosse adottata: elevato

Guida all'implementazione

Per implementare questa best practice, devi collaborare con le parti interessate presenti nell'organizzazione per concordare gli standard di comunicazione. Comunica tali standard alla tua

organizzazione. Per qualsiasi transizione IT significativa, il team di pianificazione definito può gestire con maggiore successo l'impatto del cambiamento sulle persone rispetto a un'organizzazione che ignora questa procedura. La gestione del cambiamento può essere impegnativa per le organizzazioni poiché richiede un forte consenso sulla nuova strategia di tutti i singoli collaboratori. In assenza di un team di pianificazione della transizione, la leadership ha il 100% della responsabilità di condurre comunicazioni efficaci. Quando si crea un team di pianificazione della transizione, comunica ai membri del team di collaborare con tutta la leadership organizzativa per definire e gestire comunicazioni efficaci a tutti i livelli.

Esempio del cliente

AnyCompany Retail ha sottoscritto AWS Enterprise Support e dipende da altri provider di terze parti per le sue operazioni cloud. L'azienda utilizza chat e chatop come principale mezzo di comunicazione per le attività operative. Allarmi e altre informazioni caratterizzano canali specifici. Quando qualcuno deve intervenire, il risultato desiderato viene definito in modo chiaro e, in molti casi, la persona riceve un runbook o un playbook da usare. Viene utilizzato un calendario delle modifiche per pianificare i cambiamenti più importanti ai sistemi di produzione.

Passaggi dell'implementazione

1. Crea un team principale all'interno dell'organizzazione che abbia la responsabilità di elaborare e avviare i piani di comunicazione dei cambiamenti che avvengono a più livelli all'interno dell'organizzazione.
2. Istituisci la proprietà a thread singolo per la supervisione. Offri ai singoli team la capacità di innovare in modo indipendente e bilanciare l'uso di meccanismi coerenti, consentendo così il giusto livello di ispezione e visione della direzione.
3. Collabora con le parti interessate di tutta l'organizzazione per concordare standard, procedure e piani di comunicazione.
4. Verifica che il team di comunicazione principale collabori con la leadership dell'organizzazione e del programma per creare messaggi per il personale appropriato per conto dei leader.
5. Sviluppa meccanismi di comunicazione strategici per gestire il cambiamento attraverso annunci, calendari condivisi, riunioni con tutti i partecipanti e one-on-one metodi di comunicazione di persona in modo che i membri del team abbiano aspettative adeguate sulle azioni da intraprendere.
6. Quando possibile, comunica contesto, dettagli e tempo necessari per determinare se è richiesta un'azione. Quando è necessaria un'azione, indica l'azione richiesta e il suo impatto.

7. Implementa strumenti che agevolino le comunicazioni tattiche, come chat interna, e-mail e gestione delle conoscenze.
8. Implementa meccanismi per misurare e verificare che tutte le comunicazioni portino ai risultati desiderati.
9. Stabilisci un ciclo di feedback che misuri l'efficacia delle comunicazioni, specialmente quando sono correlate alla resistenza ai cambiamenti nell'organizzazione.
10. Per tutti Account AWS, stabilisci contatti alternativi per la fatturazione, la sicurezza e le operazioni. Idealmente, ogni contatto deve essere una distribuzione di e-mail anziché una comunicazione individuale specifica.
11. Stabilisci un piano di comunicazione basato sull'escalation e sull'escalation inversa per interagire con i team interni ed esterni, compresi AWS i fornitori di assistenza e altri fornitori terzi.
12. Avvia ed esegui le strategie di comunicazione in modo coerente per tutta la durata di ciascun programma di trasformazione.
13. Assegna le priorità alle azioni ripetibili, ove possibile, per automatizzarle in sicurezza su larga scala.
14. Quando le comunicazioni sono richieste in scenari con azioni automatizzate, lo scopo della comunicazione deve essere informare i team, per il controllo o una parte del processo di gestione delle modifiche.
15. Analizza le comunicazioni provenienti dai sistemi di avviso per individuare i falsi positivi o gli avvisi creati costantemente. Rimuovi o modifica questi avvisi in modo che vengano inviati quando è richiesto l'intervento umano. Se viene attivato un avviso, fornisci un runbook o un playbook.
    a. Puoi affidarti ai documenti di AWS Systems Manager per creare playbook e runbook per gli avvisi.
16. Sono stati attivati meccanismi per fornire tempestivamente notifiche in merito ai rischi o agli eventi pianificati in modo chiaro e fruibile al fine di consentire risposte appropriate. Usa elenchi di indirizzi e-mail o canali di chat per inviare le notifiche di preavviso rispetto agli eventi pianificati.
    a. Puoi usare AWS Chatbot per inviare avvisi e rispondere agli eventi all'interno della piattaforma di messaggistica della tua organizzazione.
17. Fornisci una fonte di informazioni accessibile dove è possibile individuare gli eventi pianificati. Fornisci le notifiche degli eventi pianificati dallo stesso sistema.
    a. AWS Il calendario delle modifiche di Systems Manager può essere utilizzato per creare finestre di modifica quando possono verificarsi modifiche. In questo modo i membri del team ricevono un preavviso su quando poter effettuare la modifica in modo sicuro.

18. Monitora le notifiche di vulnerabilità e le informazioni sulle patch per capire le vulnerabilità in circolazione e i rischi potenziali associati ai componenti del tuo carico di lavoro. Invia notifiche ai membri del team in modo che possano intervenire.
    a. Puoi iscriverti ai bollettini sulla sicurezza AWS per ricevere notifiche relative a vulnerabilità su AWS.
19. Cerca opinioni e prospettive diverse: incoraggia la condivisione dei contributi da parte di tutti. Offri opportunità di comunicazione ai gruppi sottorappresentati. Distribuisci a rotazione i ruoli e le responsabilità nelle riunioni.
    a. Amplia ruoli e responsabilità: offri ai membri del team l'opportunità di assumere ruoli che altrimenti potrebbero altrimenti non ricoprire mai. Ciò consentirà loro di acquisire esperienza e nuove prospettive grazie anche alle interazioni con i nuovi membri del team, con i quali potrebbero non interagire altrimenti. Un mutuo scambio di esperienze e punti di vista vantaggioso per tutti. Con l'aumento delle prospettive, identifica le opportunità aziendali emergenti o le nuove opportunità di miglioramento. Fai in modo che i membri di un team svolgano a turno attività comuni eseguite normalmente da altri affinché comprendano richieste e impatto delle loro prestazioni.
    b. Garantisci un ambiente sicuro e ospitale: adotta policy e controlli che consentano di proteggere la sicurezza fisica e mentale dei membri del team all'interno dell'organizzazione. I membri del team devono poter interagire senza alcun timore. Quando i membri del team si sentono al sicuro e ben accolti, è più probabile che siano coinvolti e produttivi. Più è diversificata la tua organizzazione, migliore sarà la comprensione nei confronti delle persone supportate, compresi i clienti. Quando i membri del team si sentono a loro agio, sono liberi di parlare e sono sicuri di essere ascoltati, con maggiori probabilità condivideranno approfondimenti preziosi (ad esempio, opportunità di marketing, esigenze di accessibilità, segmenti di mercato non serviti, rischi non riconosciuti nel tuo ambiente).
    c. Consenti la totale partecipazione dei membri del team: fornisci le risorse necessarie ai dipendenti affinché partecipino appieno a tutte le attività correlate al lavoro. I membri del team che affrontano sfide quotidiane hanno sviluppato competenze per superarle. Queste competenze esclusive possono offrire vantaggi significativi alla tua organizzazione. Grazie al supporto di strutture adeguate, i membri del team possono apportare contributi vantaggiosi.

Risorse

Best practice correlate:

- OPS03-BP01 Fornire la sponsorizzazione esecutiva

- OPS07-BP03 Usa i runbook per eseguire le procedure
- OPS07-BP04 Usa i playbook per analizzare i problemi

Documenti correlati:

- Post sul blog AWS | Accountability and empowerment are key to high-performing agile organizations
- AWS Executive Insights | Learn to scale innovation, not complexity | Single-threaded Leaders
- AWS Security Bulletins
- Aprire CVE
- AWS Support App in Slack per gestire i casi di supporto
- Gestisci AWS le risorse nei tuoi canali Slack con AWS Chatbot

Esempi correlati:

- Well-Architected Labs: gestione di inventario e patch (Livello 100)

Servizi correlati:

- AWS Chatbot
- AWS Calendario delle modifiche di Systems Manager
- AWS Documenti Systems Manager

OPS03-BP05 La sperimentazione è incoraggiata

La sperimentazione è un catalizzatore per trasformare nuove idee in prodotti e funzionalità. La sperimentazione accelera l'apprendimento e mantiene acceso l'interesse e il coinvolgimento dei membri del team. I membri del team sono incoraggiati a sperimentare spesso per promuovere l'innovazione. Anche quando si verifica un risultato indesiderato, è comunque utile sapere quello che non bisogna fare. I membri del team non vengono puniti per gli esperimenti riusciti con risultati indesiderati.

Risultato desiderato:

- La tua organizzazione incoraggia la sperimentazione per promuovere l'innovazione.
- Gli esperimenti sono utilizzati come un'opportunità per imparare.

Anti-pattern comuni:

- Vuoi eseguire un test A/B, ma non esiste un meccanismo per eseguire l'esperimento. Distribuisci una modifica all'interfaccia utente senza la possibilità di testarla. Questo comporta un'esperienza cliente negativa.
- La tua azienda ha solo un ambiente di test e uno di produzione. Non esiste un ambiente di sperimentazione (sandbox) in cui provare nuove funzionalità o prodotti, per cui le sperimentazioni avvengono all'interno dell'ambiente di produzione.

Vantaggi dell'adozione di questa best practice:

- La sperimentazione incoraggia l'innovazione.
- Grazie alla sperimentazione puoi reagire più velocemente al feedback degli utenti.
- La tua organizzazione sviluppa una cultura dell'apprendimento.

Livello di rischio associato se questa best practice non fosse adottata: medio

Guida all'implementazione

Le sperimentazioni vanno eseguite in modo sicuro. Sfrutta più ambienti per sperimentare senza mettere a rischio le risorse di produzione. Usa il test A/B e le flag delle funzionalità per testare gli esperimenti. Offri ai membri del team la possibilità di eseguire esperimenti in un ambiente di sperimentazione (sandbox).

Esempio del cliente

AnyCompany La vendita al dettaglio incoraggia la sperimentazione. I membri del team possono dedicare il 20% della propria settimana lavorativa alla sperimentazione o all'apprendimento di nuove tecnologie. Hanno a disposizione un ambiente di sperimentazione (sandbox) in cui possono innovare. Il test A/B viene utilizzato per nuove funzionalità che possono essere così convalidate con il feedback di utenti reali.

Passaggi dell'implementazione

1. Collabora con la direzione della tua organizzazione per supportare la sperimentazione. I membri del team devono essere incoraggiati a eseguire esperimenti in modo sicuro.
2. Offri ai membri del team un ambiente in cui possono sperimentare in modo sicuro (devono avere accesso a un ambiente simile alla produzione).

   a. Puoi usarne uno separato Account AWS per creare un ambiente sandbox per la sperimentazione. AWS Control Towerpuò essere utilizzato per effettuare il provisioning di questi account.
3. Usa flag delle funzionalità e test A/B per sperimentare in modo sicuro e raccogliere il feedback degli utenti.
   a. AWS AppConfig Feature Flags offre la possibilità di creare flag di funzionalità.
   b. Amazon CloudWatch Evidently può essere utilizzato per eseguire test A/B su una distribuzione limitata.
   c. Puoi utilizzare le versioni AWS Lambda per implementare una nuova versione di una funzione per il beta testing.

Livello di impegno per il piano di implementazione: elevato. Offrire ai membri del team un ambiente in cui sperimentare in modo sicuro può richiedere investimenti significativi. Potresti anche aver bisogno di modificare il codice dell'applicazione per usare flag di funzionalità o supportare il test A/B.

Risorse

Best practice correlate:

- OPS11-BP02 Eseguire l'analisi post-incidente: imparare dagli incidenti è un fattore importante di innovazione e sperimentazione.
- OPS11-BP03 Implementazione di circuiti di feedback: i cicli di feedback costituiscono una parte importante della sperimentazione.

Documenti correlati:

- An Inside Look at the Amazon Culture: Experimentation, Failure, and Customer Obsession
- Le migliori pratiche per creare e gestire account sandbox in AWS
- Create a Culture of Experimentation Enabled by the Cloud
- Promuovere la sperimentazione e l'innovazione nel cloud presso SulAm érica Seguros
- Experiment More, Fail Less
- Organizzazione AWS dell'ambiente utilizzando più account - Sandbox OU
- Utilizzo dei flag AWS AppConfig di funzionalità

Video correlati:

- AWS On Air ft. Amazon CloudWatch Evidently | Eventi AWS
- AWS Su Air San Fran Summit 2022 ft. AWS AppConfig Integrazione di Feature Flags con Jira
- AWS re:Invent 2022 - Una distribuzione non è un rilascio: controlla i tuoi lanci con i flag di funzionalità (05-R) BOA3
- Crea Account AWS un file con a livello di codice AWS Control Tower
- Configura un AWS ambiente multi-account che utilizzi le migliori pratiche per AWS Organizations

Esempi correlati:

- AWS Sandbox per l'innovazione
- End-to-endPersonalizzazione 101 per l'e-commerce

Servizi correlati:

- Amazon CloudWatch evidentemente
- AWS AppConfig
- AWS Control Tower

OPS03-BP06 I membri del team sono incoraggiati a mantenere e accrescere le proprie competenze

I team devono aumentare le proprie competenze per adottare nuove tecnologie e supportare i cambiamenti in termini di domanda e responsabilità a supporto dei carichi di lavoro. L'ampliamento delle competenze nelle nuove tecnologie è spesso fonte di soddisfazione per i membri del team e supporta l'innovazione. Incoraggia i membri del team a perseguire e mantenere le certificazioni di settore in modo da convalidare e riconoscere le loro crescenti competenze. Pratica la formazione trasversale per promuovere il trasferimento di conoscenze e ridurre il rischio di impatto significativo in caso di perdita di membri del team qualificati ed esperti con competenze a livello istituzionale. Fornisci tempo strutturato dedicato per la formazione.

AWS fornisce risorse, tra cui il AWS Getting Started Resource Center, AWS i blog, i tech talk AWS online, AWS gli eventi e i webinar e i AWS Well-Architected Labs, che forniscono indicazioni, esempi e procedure dettagliate per istruire i team.

Risorse come AWS Support, (AWS re:Post, AWS Support Center) e documentazione AWS rimuovono gli ostacoli tecnici e consentono di migliorare le operazioni. Rivolgiti al Centro per ricevere assistenza con le tue domande. AWS Support AWS Support

AWS condivide anche le migliori pratiche e i modelli che abbiamo appreso attraverso l'utilizzo di AWSThe Amazon Builders' Library e un'ampia varietà di altro materiale didattico utile tramite il AWS blog e il podcast ufficiale. AWS

AWS Training e la certificazione include formazione gratuita tramite corsi digitali personalizzati, oltre a piani di apprendimento per ruolo o dominio. Puoi anche iscriverti a un corso di formazione con istruttore per supportare ulteriormente lo sviluppo delle competenze dei tuoi team. AWS

Risultato desiderato: la tua organizzazione valuta in modo costante le lacune nelle competenze e le colma con budget e investimenti strutturati. I team incoraggiano e incentivano i membri con attività di miglioramento delle competenze, come l'acquisizione delle principali certificazioni del settore. I team traggono vantaggio da programmi dedicati alla condivisione incrociata delle conoscenze lunch-and-learns, come giornate di immersione, hackathon e giornate di gioco. La tua organizzazione mantiene i propri sistemi di conoscenza up-to-date e la pertinenza per i membri del team con formazione trasversale, compresi i corsi di formazione iniziale per i nuovi assunti.

Anti-pattern comuni:

- In assenza di un programma di formazione strutturato e di un budget, i team riscontrano difficoltà nel tentativo di tenere il passo con l'evoluzione della tecnologia, il che si traduce in un aumento dell'attrito.
- Nell'ambito della migrazione a AWS, l'organizzazione dimostra lacune di competenze e una padronanza del cloud variabile tra i team. Senza un impegno per il miglioramento delle competenze, i team si ritrovano oberati di attività di gestione legacy e inefficienti dell'ambiente cloud, causando un aumento del lavoro degli operatori. Questo stato di esaurimento dei team aumenta l'insoddisfazione dei dipendenti.

Vantaggi dell'adozione di questa best practice: gli investimenti consapevoli della tua organizzazione nel miglioramento delle competenze dei propri team accelerano e scalano anche l'adozione e ottimizzazione del cloud. I programmi di formazione mirati favoriscono l'innovazione e creano capacità operative per consentire ai team di essere preparati a gestire gli eventi. I team investono consapevolmente nell'implementazione e nell'evoluzione delle best practice. Il morale dei team è alto e i membri apprezzano il contributo che offrono all'azienda.

Livello di rischio associato se questa best practice non fosse adottata: medio

Guida all'implementazione

Per adottare nuove tecnologie, promuovere l'innovazione e stare al passo con i cambiamenti in termini di domanda e responsabilità a supporto dei carichi di lavoro, investi continuamente nella crescita professionale dei team.

Passaggi dell'implementazione

1. Ricorri a programmi strutturati a sostegno del cloud: AWS Skills Guild offre formazione consultiva per aumentare la sicurezza nelle competenze cloud e promuovere una cultura della formazione continua.
2. Metti a disposizione le risorse per la formazione: metti a disposizione del tempo in modo strutturato e dedicato, accesso ai materiali di formazione, risorse di laboratorio e supporto alla partecipazione a conferenze e organizzazioni professionali che offrono opportunità di apprendimento da docenti e colleghi. Offri ai membri dei team junior la possibilità di contattare i membri dei team senior affinché questi fungano da mentori o possano mostrare loro come lavorano trasmettendo metodi e competenze consolidati. Incoraggia l'apprendimento dei contenuti non direttamente correlati al lavoro per avere una prospettiva più ampia.
3. Incoraggia l'uso di risorse tecniche esperte: sfrutta risorse come AWS re:Post per accedere a conoscenze consolidate e a una vibrante community.
4. Crea e gestisci un archivio di up-to-date conoscenze: utilizza piattaforme di condivisione delle conoscenze come wiki e runbook. Crea la tua fonte di conoscenza specialistica riutilizzabile con AWS re:Post Private per semplificare la collaborazione, migliorare la produttività e accelerare l'onboarding dei dipendenti.
5. Formazione del team e coinvolgimento tra team: pianifica le esigenze di formazione continua dei membri del tuo team. Offri loro l'opportunità di unirsi ad altri team (temporaneamente o definitivamente) per condividere competenze e best practice a beneficio dell'intera organizzazione.
6. Supporta il perseguimento e il mantenimento delle certificazioni di settore: favorisci l'acquisizione e il mantenimento da parte dei membri del tuo team di certificazioni di settore che convalidano quanto appreso e le loro conoscenze e riconoscono i loro risultati.

Livello di impegno per il piano di implementazione: elevato

Risorse

Best practice correlate:

- OPS03-BP01 Fornisci una sponsorizzazione esecutiva

- OPS11-BP04 Esegui la gestione della conoscenza

Documenti correlati:

- Whitepaper AWS | Cloud Adoption Framework: People Perspective
- Investing in continuous learning to grow your organization's future
- AWS Skills Guild
- AWS Training e certificazione
- AWS Support
- AWS Re: post
- Centro risorse per le nozioni di base AWS
- Blog AWS
- Conformità di Cloud AWS
- Documentazione AWS
- Il podcast ufficiale AWS.
- Colloqui tecnici online su AWS
- AWS Eventi e webinar
- AWS Well-Architected Labs
- Amazon Builders' Library

Video correlati:

- AWS re:Invent 2023 | Reskilling at the speed of cloud: Turning employees into entrepreneurs
- WS re:Invent 2023 | Building a culture of curiosity through gamification

OPS03-BP07 Team di risorse appropriati

Stabilisci il giusto numero di membri competenti del team e gli strumenti e le risorse per supportare le esigenze di carico di lavoro. Il sovraccarico dei membri del team aumenta il rischio di errore umano. Gli investimenti in strumenti e risorse, come l'automazione, consentono di scalare l'efficacia del team consentendogli di supportare un numero maggiore di carichi di lavoro senza richiedere capacità aggiuntiva.

Risultato desiderato:

- Hai assegnato al tuo team personale adeguato per acquisire le competenze necessarie a gestire i carichi di lavoro in conformità al tuo piano di migrazione. AWS Man mano che il team si è ampliato nel corso del progetto di migrazione, ha acquisito competenze nelle AWS tecnologie di base che l'azienda intende utilizzare per la migrazione o la modernizzazione delle applicazioni.
- Hai preparato con attenzione il piano per i membri del team per fare un uso efficiente delle risorse, sfruttando l'automazione e il flusso di lavoro. Un team più piccolo può ora gestire più infrastrutture per conto dei team di sviluppo delle applicazioni.
- Con il cambiamento delle priorità operative, qualsiasi vincolo di risorse viene identificato in modo proattivo per proteggere il successo delle iniziative aziendali.
- Le metriche che segnalano le difficoltà operative, ad esempio l'affaticamento da chiamata o il paging eccessivo, vengono esaminate per verificare che il personale non sia sovraccaricato.

Anti-pattern comuni:

- Il vostro personale non ha ancora migliorato AWS le proprie competenze man mano che state attuando il piano pluriennale di migrazione al cloud, il che rischia di sostenere i carichi di lavoro e di abbassare il morale dei dipendenti.
- L'intera organizzazione IT adotta le modalità di lavoro agili. L'azienda assegna le priorità al portafoglio di prodotti e stabilisce le metriche per le funzionalità che devono essere sviluppate per prime. Il processo agile non richiede che i team assegnino story point ai piani di lavoro. Di conseguenza, è impossibile conoscere il livello di capacità richiesto per il successivo lavoro o se le competenze giuste sono state assegnate al lavoro.
- Avete chiesto a un AWS partner di migrare i vostri carichi di lavoro e non disponete di un piano di transizione del supporto per i vostri team una volta che il partner avrà completato il progetto di migrazione. I team hanno difficoltà a supportare i carichi di lavoro in modo efficiente ed efficace.

Vantaggi dell'adozione di questa best practice: la tua organizzazione vanta membri del team con competenze adeguate a supportare i carichi di lavoro. L'allocazione delle risorse può adattarsi al cambiamento delle priorità senza influire sulle prestazioni. Il risultato è che i team sono in grado di supportare i carichi di lavoro, massimizzando al contempo il tempo per concentrarsi sull'innovazione per i clienti e aumentando a sua volta la soddisfazione dei dipendenti.

Livello di rischio associato se questa best practice non fosse adottata: medio

Guida all'implementazione

La pianificazione delle risorse per la migrazione al cloud deve avvenire a un livello organizzativo in linea con il piano di migrazione e l'implementazione del modello operativo desiderato per supportare il nuovo ambiente cloud. Ciò deve includere la comprensione delle tecnologie cloud utilizzate per i team di sviluppo aziendale e delle applicazioni. La leadership dell'infrastruttura e delle operazioni deve pianificare l'analisi del divario delle competenze, la formazione e la definizione dei ruoli per gli ingegneri che guidano l'adozione del cloud.

Passaggi dell'implementazione

1. Definisci i criteri per il successo dei team con metriche operative pertinenti, come la produttività del personale (ad esempio, i costi di supporto di un carico di lavoro o le ore spese dall'operatore per gli incidenti).
2. Definisci i meccanismi di pianificazione e ispezione della capacità delle risorse per verificare che il giusto equilibrio di capacità qualificata sia disponibile quando necessario e possa essere modificato nel tempo.
3. Crea meccanismi, ad esempio inviando un sondaggio mensile ai team, per comprendere le sfide legate al lavoro che hanno un impatto sui team, come l'aumento delle responsabilità, i cambiamenti nella tecnologia, la mancanza di personale o l'aumento dei clienti supportati.
4. Utilizza questi meccanismi per interagire con i team e individuare le tendenze che possono contribuire alle sfide relative alla produttività dei dipendenti. Quando i team sono influenzati da fattori esterni, rivaluta gli obiettivi e modifica i target in base alle esigenze. Individua gli ostacoli che impediscono i progressi dei team.
5. Verifica con regolarità se le risorse attualmente allocate sono ancora sufficienti o se occorre aggiungere e apportare le modifiche appropriate ai team di supporto.

Livello di impegno per il piano di implementazione: medio

Risorse

Best practice correlate:

- OPS03-BP06 I membri del team sono incoraggiati a mantenere e accrescere le proprie competenze
- OPS09-BP03 Rivedi le metriche operative e dai priorità al miglioramento
- OPS10-BP01 Utilizza un processo per la gestione di eventi, incidenti e problemi

- OPS10-BP07 Automatizza le risposte agli eventi

Documenti correlati:

- Cloud AWS Framework di adozione: prospettiva delle persone
- Becoming a Future-Ready Enterprise
- Prioritize your Employees' Skills to Drive Business Growth
- Organizzazioni ad alte prestazioni: il team da due pizze Amazon
- How Cloud-Mature Enterprises Succeed

Preparazione

Questions

- OPS4. Come si implementa l'osservabilità nel carico di lavoro?
- OPS5. In che modo riduci i difetti, favorisci la correzione e migliori il flusso nella produzione?
- OPS6. In che modo mitighi i rischi dell'implementazione?
- OPS7. Come fai a sapere se hai tutto pronto per supportare un carico di lavoro?

OPS4. Come si implementa l'osservabilità nel carico di lavoro?

Implementare l'osservabilità nel carico di lavoro ti permette di comprendere lo stato di quest'ultimo e di adottare decisioni basate sui dati e che riflettono i requisiti aziendali.

Best practice

- OPS04-BP01 Identifica gli indicatori chiave di prestazione
- OPS04-BP02 Implementare la telemetria delle applicazioni
- OPS04-BP03 Implementare la telemetria dell'esperienza utente
- OPS04-BP04 Implementare la telemetria delle dipendenze
- OPS04-BP05 Implementare la tracciabilità distribuita

OPS04-BP01 Identifica gli indicatori chiave di prestazione

L'implementazione dell'osservabilità nel carico di lavoro inizia con la comprensione del suo stato e l'adozione di decisioni basate sui dati che riflettono i requisiti aziendali. Uno dei modi più efficaci per

garantire l'allineamento tra le attività di monitoraggio e gli obiettivi aziendali consiste nella definizione e nel monitoraggio degli indicatori chiave di performance (). KPIs

Risultato desiderato: pratiche di osservabilità efficienti e strettamente allineate agli obiettivi aziendali garantiscono che le attività di monitoraggio siano sempre al servizio di risultati aziendali tangibili.

Anti-pattern comuni:

- IndefinitoKPIs: lavorare senza un sistema chiaro KPIs può portare a un monitoraggio eccessivo o insufficiente e alla mancanza di segnali vitali.
- StaticoKPIs: non rivisitare o perfezionare man mano che il carico di lavoro o KPIs gli obiettivi aziendali si evolvono.
- Disallineamento: concentrarsi su metriche tecniche non direttamente correlate ai risultati aziendali o che sono più difficili da correlare ai problemi del mondo reale.

Vantaggi dell'adozione di questa best practice:

- Facilità di identificazione dei problemi: le aziende KPIs spesso evidenziano i problemi in modo più chiaro rispetto alle metriche tecniche. Un calo aziendale KPI può individuare un problema in modo più efficace rispetto all'analisi di numerose metriche tecniche.
- Allineamento aziendale: assicura che le attività di monitoraggio supportino direttamente gli obiettivi aziendali.
- Efficienza: viene data la priorità alle risorse di monitoraggio e al focus sulle metriche che contano.
- Proattività: riconoscere e risolvere i problemi prima che abbiano implicazioni aziendali più ampie.

Livello di rischio associato se questa best practice non fosse adottata: elevato

Guida all'implementazione

Per definire in modo efficace il carico di lavoro: KPIs

1. Inizia con i risultati aziendali: prima di approfondire le metriche, comprendi i risultati aziendali desiderati. È stato rilevato un aumento delle vendite, un maggiore coinvolgimento degli utenti o tempi di risposta più rapidi?
2. Correla le metriche tecniche con gli obiettivi aziendali: non tutte le metriche tecniche influiscono direttamente sui risultati aziendali. Identifica quelli che lo fanno, ma spesso è più semplice identificare un problema utilizzando un'azienda. KPI

3. Usa Amazon CloudWatch: Employ CloudWatch per definire e monitorare le metriche che rappresentano le tue. KPIs
4. Rivedi e aggiorna regolarmenteKPIs: man mano che il carico di lavoro e la tua attività si evolvono, mantieni i tuoi dati pertinenti. KPIs
5. Coinvolgi le parti interessate: coinvolgi i team tecnici e aziendali nella definizione e nella revisione. KPIs

Livello di impegno per il piano di implementazione: medio

Risorse

Best practice correlate:

- the section called “OPS04-BP02 Implementare la telemetria delle applicazioni”
- the section called “OPS04-BP03 Implementare la telemetria dell'esperienza utente”
- the section called “OPS04-BP04 Implementare la telemetria delle dipendenze”
- the section called “OPS04-BP05 Implementare la tracciabilità distribuita”

Documenti correlati:

- AWS Migliori pratiche di osservabilità
- CloudWatch Guida per l'utente
- AWS Corso Observability Skill Builder

Video correlati:

- Developing an observability strategy

Esempi correlati:

- One Observability Workshop

OPS04-BP02 Implementare la telemetria delle applicazioni

La telemetria dell'applicazione è la base su cui si fonda l'osservabilità del carico di lavoro. È fondamentale emettere dati di telemetria che offrano approfondimenti utili sullo stato dell'applicazione

e sul raggiungimento degli obiettivi sia tecnici sia aziendali. Dalla risoluzione dei problemi alla misurazione dell'impatto di una nuova funzionalità o alla garanzia dell'allineamento con gli indicatori chiave di prestazione aziendali (KPIs), la telemetria delle applicazioni influenza il modo in cui create, gestite ed evolvete il carico di lavoro.

Metriche, log e tracce costituiscono i tre pilastri principali dell'osservabilità. Questi operano come strumenti diagnostici che descrivono lo stato dell'applicazione. Nel tempo, aiutano a creare criteri di base e a identificare le anomalie. Tuttavia, per garantire l'allineamento tra le attività di monitoraggio e gli obiettivi aziendali, è fondamentale definire e monitorare. KPIs KPIsLe aziende spesso semplificano l'identificazione dei problemi rispetto alle sole metriche tecniche.

Altri tipi di telemetria, come il monitoraggio degli utenti in tempo reale (RUM) e le transazioni sintetiche, completano queste fonti di dati primarie. RUMoffre approfondimenti sulle interazioni degli utenti in tempo reale, mentre le transazioni sintetiche simulano i potenziali comportamenti degli utenti, aiutando a individuare i colli di bottiglia prima che gli utenti reali li incontrino.

Risultato desiderato: ottieni approfondimenti utili sulle prestazioni del tuo carico di lavoro. Questi approfondimenti consentono di prendere decisioni proattive sull'ottimizzazione delle prestazioni, ottenere una maggiore stabilità del carico di lavoro, semplificare i processi CI/CD e utilizzare le risorse in modo efficace.

Anti-pattern comuni:

- Osservabilità incompleta: trascurare l'incorporazione dell'osservabilità a ogni livello del carico di lavoro, con conseguenti punti ciechi che possono nascondere le prestazioni vitali del sistema e gli approfondimenti sul comportamento.
- Visualizzazione frammentata dei dati: quando i dati sono sparsi su più strumenti e sistemi, diventa difficile mantenere una visione olistica dello stato e delle prestazioni del carico di lavoro.
- Problemi segnalati dagli utenti: un segno della mancanza di un rilevamento proattivo dei problemi tramite telemetria e monitoraggio aziendale. KPI

Vantaggi dell'adozione di questa best practice:

- Processo decisionale informato: con gli approfondimenti tratti dalla telemetria e dal business, puoi prendere decisioni basate sui dati. KPIs
- Migliore efficienza operativa: l'utilizzo delle risorse basato sui dati porta a un miglioramento dell'efficienza risparmiando sui costi.

- Maggiore stabilità del carico di lavoro: rilevamento e risoluzione più rapidi dei problemi con conseguente aumento dei tempi di attività.
- Processi CI/CD semplificati: gli approfondimenti ricavati dai dati di telemetria facilitano il perfezionamento dei processi e la distribuzione affidabile del codice.

Livello di rischio associato se questa best practice non fosse adottata: elevato

Guida all'implementazione

Per implementare la telemetria delle applicazioni per il tuo carico di lavoro, utilizza servizi AWS come Amazon e. CloudWatch AWS X-Ray Amazon CloudWatch offre una suite completa di strumenti di monitoraggio che ti consentono di osservare le tue risorse e applicazioni in ambienti locali AWS e locali. Raccoglie, tiene traccia e analizza le metriche, consolida e monitora i dati di log e risponde alle modifiche che interessano le risorse, migliorando la comprensione del funzionamento del carico di lavoro. In parallelo, ti AWS X-Ray consente di tracciare, analizzare ed eseguire il debug delle tue applicazioni, offrendoti una comprensione approfondita del comportamento del tuo carico di lavoro. Grazie a funzionalità come mappe dei servizi, distribuzioni della latenza e tempistiche di tracciamento, AWS X-Ray fornisce informazioni dettagliate sulle prestazioni del carico di lavoro e sui colli di bottiglia che lo influiscono.

Passaggi dell'implementazione

1. Identifica quali dati raccogliere: definisci le metriche, i log e le tracce essenziali che potrebbero offrire importanti informazioni dettagliate sullo stato, le prestazioni e il comportamento del tuo carico di lavoro.
2. Implementa l'CloudWatchagente: l' CloudWatch agente è fondamentale nell'acquisizione dei parametri e dei log di sistema e delle applicazioni dal carico di lavoro e dall'infrastruttura sottostante. L' CloudWatch agente può essere utilizzato anche per raccogliere OpenTelemetry o inviare tracce a raggi X e inviarle a X-Ray.
3. Implementa il rilevamento delle anomalie per log e metriche: utilizza il rilevamento delle anomalie CloudWatch nei log e il rilevamento delle anomalie nelle CloudWatchmetriche per identificare automaticamente le attività insolite nelle operazioni dell'applicazione. Questi strumenti utilizzano algoritmi di machine learning per rilevare e comunicare le anomalie, migliorando le capacità di monitoraggio e accelerando i tempi di risposta a potenziali interruzioni o minacce alla sicurezza. Configura queste funzionalità per gestire in modo proattivo lo stato e la sicurezza delle applicazioni.

4. Proteggi i dati sensibili dei log: utilizza la protezione dei dati di Amazon CloudWatch Logs per mascherare le informazioni sensibili all'interno dei tuoi log. Questa funzionalità aiuta a mantenere la privacy e la conformità con il rilevamento e il mascheramento automatici dei dati sensibili prima dell'accesso. Implementa il mascheramento dei dati per gestire e proteggere in modo sicuro i dettagli sensibili come le informazioni di identificazione personale (). PII
5. Definisci e monitora il businessKPIs: stabilisci metriche personalizzate in linea con i risultati aziendali.
6. Strumenta la tua applicazione con AWS X-Ray: oltre a implementare l' CloudWatchagente, è fondamentale strumentare l'applicazione per emettere dati di traccia. Questo processo può fornire ulteriori approfondimenti sul comportamento e sulle prestazioni del carico di lavoro.
7. Standardizza la raccolta dei dati nell'applicazione: standardizza le pratiche di raccolta dei dati nell'intera applicazione. L'uniformità aiuta a correlare e analizzare i dati, fornendo una visione completa del comportamento dell'applicazione.
8. Implementa l'osservabilità tra account: migliora l'efficienza del monitoraggio su più account con l'osservabilità tra più account di Account AWS Amazon CloudWatch . Con questa funzionalità, puoi consolidare metriche, log e allarmi di diversi account in un'unica visualizzazione, semplificando la gestione e migliorando i tempi di risposta per i problemi identificati nell'ambiente dell'organizzazione. AWS
9. Analizza e agisci in base ai dati: una volta completata la raccolta e la normalizzazione dei dati, usa Amazon CloudWatch per l'analisi di metriche e log e AWS X-Rayper l'analisi delle tracce. Tale analisi può fornire approfondimenti cruciali sullo stato, le prestazioni e il comportamento del carico di lavoro, guidando il processo decisionale.

Livello di impegno per il piano di implementazione: elevato

Risorse

Best practice correlate:

- OPS04-BP01 Definisci il carico di lavoro KPIs
- OPS04-BP03 Implementare la telemetria delle attività degli utenti
- OPS04-BP04 Implementare la telemetria delle dipendenze
- OPS04-BP05 Implementare la tracciabilità delle transazioni

Documenti correlati:

- AWS Observability Best Practices
- Guida per l'utente di CloudWatch
- AWS X-Ray Guida per gli sviluppatori
- Strumentazione di sistemi distribuiti per visibilità operativa
- AWS Observability Skill Builder Course
- Cosa c'è di nuovo con Amazon CloudWatch
- Cosa c'è di nuovo con AWS X-Ray

Video correlati:

- AWS re:Invent 2022 - Le migliori pratiche di osservabilità su Amazon
- AWS re:Invent 2022 - Sviluppo di una strategia di osservabilità

Esempi correlati:

- One Observability Workshop
- AWS Libreria di soluzioni: monitoraggio delle applicazioni con Amazon CloudWatch

OPS04-BP03 Implementare la telemetria dell'esperienza utente

Acquisire informazioni approfondite sulle esperienze dei clienti e sulle interazioni con la tua applicazione è fondamentale. Il monitoraggio degli utenti reali (RUM) e le transazioni sintetiche sono strumenti potenti per questo scopo. RUMfornisce dati sulle interazioni reali degli utenti garantendo una prospettiva non filtrata della soddisfazione degli utenti, mentre le transazioni sintetiche simulano le interazioni degli utenti, aiutando a rilevare potenziali problemi ancor prima che abbiano un impatto sugli utenti reali.

Risultato desiderato: una visione olistica dell'esperienza del cliente, il rilevamento proattivo dei problemi e l'ottimizzazione delle interazioni degli utenti per offrire esperienze digitali fluide.

Anti-pattern comuni:

- Applicazioni senza monitoraggio reale degli utenti (): RUM
  - Rilevamento ritardato dei problemi: in caso contrarioRUM, potreste non accorgervi dei rallentamenti o dei problemi di prestazioni fino a quando gli utenti non si lamentano. Questo approccio reattivo può causare insoddisfazione nei clienti.

    - Mancanza di informazioni sull'esperienza utente: non utilizzarla RUM significa perdere dati cruciali che mostrano come gli utenti reali interagiscono con l'applicazione, limitando la capacità di ottimizzare l'esperienza utente.
- Applicazioni senza transazioni sintetiche:
    - Casi limite trascurati: le transazioni sintetiche consentono di testare percorsi e funzioni che potrebbero non essere utilizzati frequentemente dagli utenti tipici, ma che sono fondamentali per determinate funzioni aziendali. Senza di esse, questi percorsi potrebbero non funzionare correttamente e passare inosservati.
    - Verifica della presenza di problemi quando l'applicazione non viene utilizzata: i test sintetici regolari possono simulare situazioni in cui gli utenti reali non interagiscono attivamente con l'applicazione, garantendo che il sistema funzioni sempre correttamente.

Vantaggi dell'adozione di questa best practice:

- Rilevamento proattivo dei problemi: identifica e risolvi i problemi potenziali prima che abbiano un impatto sugli utenti reali.
- Esperienza utente ottimizzata: il feedback continuo fornito RUM aiuta a perfezionare e migliorare l'esperienza utente complessiva.
- Informazioni approfondite sulle prestazioni del dispositivo e del browser: scopri come si comporta la tua applicazione in vari dispositivi e browser e implementa ulteriori ottimizzazioni.
- Flussi di lavoro aziendali convalidati: transazioni sintetiche regolari assicurano che le funzionalità principali e i percorsi critici siano operativi ed efficienti in maniera costante.
- Prestazioni delle applicazioni migliorate: sfrutta le informazioni approfondite raccolte dai dati degli utenti reali per migliorare la reattività e l'affidabilità delle applicazioni.

Livello di rischio associato se questa best practice non fosse adottata: elevato

Guida all'implementazione

Per sfruttare RUM e sintetizzare le transazioni per la telemetria delle attività degli utenti, offre AWS servizi come Amazon e Amazon CloudWatch RUM Synthetics. CloudWatch Metriche, log e tracce, insieme ai dati sulle attività degli utenti, forniscono una visione completa dello stato operativo dell'applicazione e dell'esperienza utente.

Passaggi dell'implementazione

1. Implementa Amazon CloudWatch RUM: integra la tua applicazione con CloudWatch RUM per raccogliere, analizzare e presentare dati utente reali.
   a. Usa la CloudWatch RUM JavaScript libreria per l'integrazione RUM con la tua applicazione.
   b. Configura pannelli di controllo per visualizzare e monitorare i dati relativi agli utenti reali.
2. Configura CloudWatch Synthetics: crea canaries, o routine con script, che simulano le interazioni degli utenti con la tua applicazione.
   a. Definisci i flussi di lavoro e i percorsi critici delle applicazioni.
   b. Progetta canarini utilizzando gli script CloudWatch Synthetics per simulare le interazioni degli utenti per questi percorsi.
   c. Pianifica e monitora i canary affinché si attivino a intervalli specifici, in modo da garantire controlli costanti delle prestazioni.
3. Analizza e agisci in base ai dati: utilizza i dati e le transazioni sintetiche per ottenere informazioni RUM e adottare misure correttive quando vengono rilevate anomalie. Utilizza CloudWatch dashboard e allarmi per rimanere informato.

Livello di impegno per il piano di implementazione: medio

Risorse

Best practice correlate:

- OPS04-BP01 Identifica gli indicatori chiave di prestazione
- OPS04-BP02 Implementare la telemetria delle applicazioni
- OPS04-BP04 Implementare la telemetria delle dipendenze
- OPS04-BP05 Implementare la tracciabilità distribuita

Documenti correlati:

- CloudWatch RUMGuida Amazon
- Guida Amazon CloudWatch Synthetics

Video correlati:

- Ottimizza le applicazioni attraverso approfondimenti sugli utenti finali con Amazon CloudWatch RUM
- AWS su Air ft. Monitoraggio degli utenti in tempo reale per Amazon CloudWatch

Esempi correlati:

- One Observability Workshop
- Repository Git per Amazon CloudWatch RUM Web Client
- Utilizzo di Amazon CloudWatch Synthetics per misurare il tempo di caricamento delle pagine

OPS04-BP04 Implementare la telemetria delle dipendenze

La telemetria delle dipendenze è essenziale per monitorare lo stato e le prestazioni dei servizi e dei componenti esterni su cui si basa il carico di lavoro. Fornisce informazioni preziose sulla raggiungibilità, sui timeout e su altri eventi critici relativi a dipendenze, ad esempio database o terze parti. DNS APIs Dotando l'applicazione di strumenti per generare metriche, log e tracce relative a queste dipendenze, acquisisci una comprensione più chiara dei potenziali colli di bottiglia, problemi di prestazioni o errori che potrebbero influire sul carico di lavoro.

Risultato desiderato: le dipendenze su cui si basa il carico di lavoro funzionano come previsto, consentendo di gestire i problemi in modo proattivo e garantendo prestazioni ottimali del carico di lavoro.

Anti-pattern comuni:

- Scarsa attenzione alle dipendenze esterne: il focus è rivolto esclusivamente alle metriche interne dell'applicazione, trascurando quelle legate alle dipendenze esterne.
- Mancanza di monitoraggio proattivo: si attende che si verifichino problemi anziché monitorare costantemente lo stato e le prestazioni delle dipendenze.
- Monitoraggio isolato in comparti: utilizzo di strumenti di monitoraggio multipli ed eterogenei che possono portare a visioni dello stato delle dipendenze frammentate e incoerenti.

Vantaggi dell'adozione di questa best practice:

- Maggiore affidabilità del carico di lavoro: viene garantito che le dipendenze esterne siano costantemente disponibili e funzionino in modo ottimale.

- Rilevamento e risoluzione dei problemi più rapidi: identificazione e risoluzione proattiva dei problemi relativi alle dipendenze prima che influiscano sul carico di lavoro.
- Visione completa: acquisizione di una visione olistica dei componenti interni ed esterni che influenzano lo stato del carico di lavoro.
- Scalabilità del carico di lavoro migliorata: grazie alla comprensione dei limiti di scalabilità e delle caratteristiche prestazionali delle dipendenze esterne.

Livello di rischio associato se questa best practice non fosse adottata: elevato

Guida all'implementazione

Implementa la telemetria delle dipendenze iniziando con l'identificazione dei servizi, dell'infrastruttura e dei processi da cui dipende il carico di lavoro. Esegui una valutazione quantitativa delle condizioni ottimali nelle quali tali dipendenze funzionano come previsto e poi determina quali dati sono necessari per misurarle. Con queste informazioni, puoi creare dashboard e avvisi che forniscono approfondimenti ai tuoi team operativi sullo stato di tali dipendenze. Utilizza AWS gli strumenti per scoprire e quantificare gli impatti quando le dipendenze non riescono a soddisfare le esigenze. Riesamina costantemente la tua strategia per tenere conto dei cambiamenti relativi a priorità, obiettivi e alle informazioni dettagliate acquisite.

Passaggi dell'implementazione

Per implementare efficacemente la telemetria delle dipendenze:

1. Identifica le dipendenze esterne: collabora con le parti interessate per individuare le dipendenze esterne sulle quali si basa il tuo carico di lavoro. Le dipendenze esterne possono comprendere servizi come database esterni, percorsi di connettività di rete di terze parti APIs verso altri ambienti e servizi. DNS Il primo passo verso un'efficace telemetria delle dipendenze è acquisire una comprensione totale di quali esse siano.
2. Sviluppa una strategia di monitoraggio: una volta acquisito un quadro chiaro delle dipendenze esterne, progetta una strategia di monitoraggio ad hoc per esse. Ciò implica la comprensione della criticità di ogni dipendenza, del suo comportamento previsto e degli eventuali accordi o obiettivi sui livelli di servizio associati (o). SLA SLTs Imposta avvisi proattivi che ti informino riguardo a cambiamenti di stato o deviazioni delle prestazioni.
3. Usa il monitoraggio della rete: utilizza Internet Monitor e Network Monitor per informazioni complete sulle condizioni globali di Internet e della rete. Questi strumenti consentono di comprendere e rispondere alle interruzioni, ai malfunzionamenti o al degrado delle prestazioni che influiscono sulle dipendenze esterne.

4. Resta informato AWS Health Dashboard: fornisce avvisi e indicazioni per la risoluzione di eventuali eventi che possono AWS influire sui servizi.
   a. Monitora AWS Health gli eventi con EventBridge le regole di Amazon o esegui l'integrazione programmatica con AWS Health API per automatizzare le azioni quando ricevi AWS Health eventi. Può trattarsi di azioni generali, come l'invio di tutti i messaggi pianificati sugli eventi del ciclo di vita a un'interfaccia di chat, oppure azioni specifiche, come l'avvio di un flusso di lavoro in uno strumento di gestione dei servizi IT.
   b. Se lo utilizzi AWS Organizations, aggrega AWS Health gli eventi tra gli account.
5. Strumenta la tua applicazione con AWS X-Ray: AWS X-Ray fornisce informazioni dettagliate sulle prestazioni delle applicazioni e sulle relative dipendenze sottostanti. La tracciatura delle richieste dall'inizio alla fine ti permette di identificare colli di bottiglia o guasti nei servizi o nei componenti esterni su cui si basa l'applicazione.
6. Usa Amazon DevOps Guru: questo servizio basato sull'apprendimento automatico identifica i problemi operativi, prevede quando potrebbero verificarsi problemi critici e consiglia azioni specifiche da intraprendere. Fornisce un supporto prezioso per acquisire approfondimenti sulle dipendenze e assicurarsi che queste non siano la fonte di problemi operativi.
7. Monitora regolarmente: monitora le metriche e i log relativi alle dipendenze esterne in maniera costante. Imposta avvisi per comportamenti imprevisti o prestazioni ridotte.
8. Convalida dopo le modifiche: ogni volta che una dipendenza esterna è interessata da un aggiornamento o una modifica, convalidane le prestazioni e verifica che queste siano in linea con i requisiti dell'applicazione.

Livello di impegno per il piano di implementazione: medio

Risorse

Best practice correlate:

- OPS04-BP01 Definisci il carico di lavoro KPIs
- OPS04-BP02 Implementare la telemetria delle applicazioni
- OPS04-BP03 Implementare la telemetria delle attività degli utenti
- OPS04-BP05 Implementare la tracciabilità delle transazioni
- OPS08-BP04 Creare avvisi fruibili

Documenti correlati:

- Guida per AWS Health Dashboard l'utente di Amazon Personal
- AWS Internet Monitor User Guide
- AWS X-Ray Guida per gli sviluppatori
- AWS DevOpsGuida per l'utente Guru

Video correlati:

- Visibility into how internet issues impact app performance
- Introduzione ad Amazon DevOps Guru
- Gestisci gli eventi del ciclo di vita delle risorse su larga scala con AWS Health

Esempi correlati:

- Ottenere informazioni operative AIOps con Amazon DevOps Guru
- AWS Health Consapevole
- Utilizzo del filtro basato su tag per gestire il AWS Health monitoraggio e gli avvisi su larga scala

OPS04-BP05 Implementare la tracciabilità distribuita

Il tracciamento distribuito offre un modo per monitorare e visualizzare le richieste mentre attraversano vari componenti di un sistema distribuito. Acquisendo i dati di tracciamento da più fonti e analizzandoli in una vista unificata, i team possono comprendere meglio il flusso delle richieste, in quali punti sono presenti colli di bottiglia e dove devono concentrare gli sforzi di ottimizzazione.

Risultato desiderato: una visione olistica del flusso delle richieste nel tuo sistema distribuito, che ti permette di ottenere un debug preciso, prestazioni ottimizzate e migliori esperienze utente.

Anti-pattern comuni:

- Strumentazione incoerente: non tutti i servizi in un sistema distribuito sono dotati di strumentazione per il monitoraggio.
- Ignorare la latenza: concentrarsi solo sugli errori e non considerare la latenza o il graduale deterioramento delle prestazioni.

Vantaggi dell'adozione di questa best practice:

- Panoramica completa del sistema: visualizzazione dell'intero percorso delle richieste, dall'ingresso all'uscita.
- Debug avanzato: identificazione rapida dei punti in cui si verificano guasti o problemi di prestazioni.
- Esperienza utente migliorata: monitoraggio e ottimizzazione in base ai dati effettivi dell'utente, garantendo che il sistema soddisfi le esigenze del mondo reale.

Livello di rischio associato se questa best practice non fosse adottata: elevato

Guida all'implementazione

Inizia identificando tutti gli elementi del carico di lavoro che richiedono strumentazione. Una volta presi in considerazione tutti i componenti, sfrutta strumenti come AWS X-Ray e OpenTelemetry per raccogliere dati di traccia per l'analisi con strumenti come X-Ray e Amazon Map. CloudWatch ServiceLens Partecipa a revisioni periodiche con gli sviluppatori e integra queste discussioni con strumenti come Amazon DevOps Guru, X-Ray Analytics e X-Ray Insights per aiutarti a scoprire risultati più approfonditi. Imposta avvisi basati sui dati di tracciamento per notificare quando i risultati sono a rischio, come definito nel piano di monitoraggio del carico di lavoro.

Passaggi dell'implementazione

Per implementare il tracciamento distribuito in modo efficace:

1. Adotta AWS X-Ray: implementa X-Ray nella tua applicazione per ottenere informazioni dettagliate sul suo comportamento, comprenderne le prestazioni e individuare i punti critici. Utilizza X-Ray Insights per l'analisi automatica dei tracciamenti.
2. Strumenta i tuoi servizi: verifica che ogni servizio, da una AWS Lambdafunzione a un'EC2istanza, invii dati di traccia. Maggiore è il numero di servizi che offri, più chiara è la end-to-end visione.
3. Incorpora il monitoraggio degli utenti CloudWatch reali e il monitoraggio sintetico: integra il monitoraggio degli utenti reali (RUM) e il monitoraggio sintetico con X-Ray. Ciò ti consente di acquisire esperienze utenti del mondo reale e simulare le interazioni degli utenti per identificare potenziali problemi.
4. Usa l'CloudWatch agente: l'agente può inviare tracce da raggi X o OpenTelemetry, migliorando la profondità delle informazioni ottenute.
5. Usa Amazon DevOps Guru: DevOps Guru utilizza i dati di X-Ray, CloudWatch AWS Config, e AWS CloudTrail per fornire consigli pratici.
6. Analizza le tracce: esamina regolarmente i dati di tracciamento per individuare schemi, anomalie o colli di bottiglia che possono influire sulle prestazioni dell'applicazione.

7. Imposta avvisi: configura gli allarmi per schemi insoliti o latenze prolungate, CloudWatchper una risoluzione proattiva dei problemi.
8. Miglioramento continuo: riesamina la tua strategia di tracciamento man mano che aggiungi o modifichi servizi per acquisire tutti i punti dati pertinenti.

Livello di impegno per il piano di implementazione: medio

Risorse

Best practice correlate:

- OPS04-BP01 Identifica gli indicatori chiave di prestazione
- OPS04-BP02 Implementare la telemetria delle applicazioni
- OPS04-BP03 Implementare la telemetria dell'esperienza utente
- OPS04-BP04 Implementare la telemetria delle dipendenze

Documenti correlati:

- AWS X-Ray Guida per gli sviluppatori
- Guida per CloudWatch l'utente dell'agente Amazon
- Guida per l'utente di Amazon DevOps Guru

Video correlati:

- Usa Insights AWS X-Ray
- AWS su Air ft. Osservabilità: Amazon CloudWatch e AWS X-Ray

Esempi correlati:

- Strumentazione della tua applicazione per AWS X-Ray

OPS5. In che modo riduci i difetti, favorisci la correzione e migliori il flusso nella produzione?

Adotta approcci che migliorino il flusso delle modifiche nella produzione, che attivino la rifattorizzazione e il feedback veloce su qualità e correzione di errori. Tali approcci accelerano

l'ingresso in produzione delle modifiche vantaggiose, contengono i problemi che si sono diffusi e permettono di ottenere una rapida identificazione e risoluzione dei problemi introdotti attraverso le attività di implementazione.

Best practice

- OPS05-BP01 Usa il controllo della versione
- OPS05-BP02 Testare e convalidare le modifiche
- OPS05-BP03 Utilizzare sistemi di gestione della configurazione
- OPS05-BP04 Usa sistemi di gestione della compilazione e dell'implementazione
- OPS05-BP05 Eseguire la gestione delle patch
- OPS05-BP06 Condividi gli standard di progettazione
- OPS05-BP07 Implementare pratiche per migliorare la qualità del codice
- OPS05-BP08 Usa più ambienti
- OPS05-BP09 Apporta modifiche frequenti, piccole e reversibili
- OPS05- BP1 0 Integrazione e implementazione completamente automatizzate

OPS05-BP01 Usa il controllo della versione

Utilizza il controllo delle versioni per attivare il monitoraggio di modifiche e rilasci.

Molti AWS servizi offrono funzionalità di controllo della versione. Utilizza un sistema di revisione o di controllo del codice sorgente, come esempio AWS CodeCommit, per la gestione di codice e altri artefatti, come i modelli AWS CloudFormation con controllo delle versioni della tua infrastruttura.

Risultato desiderato: collaborazione dei team nell'ambito del codice. Una volta unito, il codice è coerente e nessuna modifica viene persa. Gli errori possono essere facilmente ripristinati mediante il corretto controllo delle versioni.

Anti-pattern comuni:

- Hai sviluppato e archiviato il codice sulla workstation. Si è verificato un errore di archiviazione non recuperabile sulla workstation e il codice è andato perso.
- Dopo aver sovrascritto il codice esistente con le modifiche, riavvii l'applicazione e non è più utilizzabile. Non è possibile ripristinare la modifica.
- Hai un blocco di scrittura su un file di report che deve essere modificato da altri utenti. Ti contattano per chiederti di smettere di utilizzarlo in modo che possano completare le loro attività.

- Il team di ricerca ha lavorato a un'analisi dettagliata che definisce il tuo lavoro futuro. Qualcuno ha salvato accidentalmente la lista della spesa nel report finale. Non puoi ripristinare la modifica e devi ricreare il report.

Vantaggi dell'adozione di questa best practice: grazie alle funzionalità di controllo delle versioni, puoi ripristinare facilmente gli stati validi noti e le versioni precedenti e limitare il rischio di perdita degli asset.

Livello di rischio associato se questa best practice non fosse adottata: elevato

Guida all'implementazione

Mantieni gli asset in repository con controllo delle versioni. In questo modo si supporta il monitoraggio delle modifiche, l'implementazione di nuove versioni, il rilevamento delle modifiche apportate alle versioni esistenti e il ripristino delle versioni precedenti, ad esempio il rollback a uno stato corretto noto in caso di errore. Integra nelle tue procedure le funzionalità di controllo delle versioni dei sistemi di gestione delle configurazioni.

Risorse

Best practice correlate:

- OPS05-BP04 Usa sistemi di gestione della compilazione e dell'implementazione

Documenti correlati:

- Che cos'è AWS CodeCommit?

Video correlati:

- Introduzione a AWS CodeCommit

OPS05-BP02 Testare e convalidare le modifiche

Ogni modifica apportata deve essere testata per evitare errori in produzione. Questa best practice si concentra sulla verifica delle modifiche dal controllo di versione alla creazione dell'artefatto. Oltre alle modifiche al codice dell'applicazione, i test dovrebbero includere l'infrastruttura, la configurazione, i controlli di sicurezza e le procedure operative. I test assumono molte forme, dai test unitari all'analisi

dei componenti software (SCA). Spostando i test più a sinistra nel processo di integrazione e consegna del software ottieni una maggiore certezza della qualità degli artefatti.

L'organizzazione deve sviluppare standard di test per tutti gli artefatti software. I test automatizzati riducono la fatica ed evitano gli errori dei test manuali. I test manuali potrebbero essere necessari in alcuni casi. Gli sviluppatori devono avere accesso ai risultati dei test automatizzati per creare cicli di feedback che migliorino la qualità del software.

Risultato desiderato: le modifiche software vengono testate prima del rilascio. Gli sviluppatori hanno accesso ai risultati dei test e alle convalide. La tua organizzazione ha uno standard per i test che applica a tutte le modifiche software.

Anti-pattern comuni:

- Implementi una nuova modifica software senza test. Non funziona in produzione e genera un'interruzione.
- I nuovi gruppi di sicurezza vengono implementati AWS CloudFormation senza essere testati in un ambiente di preproduzione. I gruppi di sicurezza rendono la tua app irraggiungibile per i clienti.
- Un metodo viene modificato, ma non ci sono test di unità. Il software ha esito negativo quando viene implementato in produzione.

Vantaggi dell'adozione di questa best practice: riduzione della percentuale di errori di modifica delle implementazioni software. La qualità del software viene migliorata. Gli sviluppatori hanno una maggiore consapevolezza della fattibilità del loro codice. Le policy di sicurezza possono essere implementate in maniera affidabile per supportare la conformità dell'organizzazione. Le modifiche all'infrastruttura, come gli aggiornamenti automatici delle policy di dimensionamento, vengono testate in anticipo per soddisfare le esigenze del traffico.

Livello di rischio associato se questa best practice non fosse adottata: elevato

Guida all'implementazione

I test vengono eseguiti su tutte le modifiche, dal codice dell'applicazione all'infrastruttura, come parte della pratica di integrazione continua. I risultati dei test vengono pubblicati in modo che gli sviluppatori abbiano un feedback rapido. La tua organizzazione ha uno standard per i test che applica a tutte le modifiche software.

Usa la potenza dell'IA generativa con Amazon Q Developer per migliorare la produttività degli sviluppatori e la qualità del codice. Amazon Q Developer include la generazione di suggerimenti di

codice (basati su modelli linguistici di grandi dimensioni), la produzione di test di unità (comprese le condizioni limite) e il miglioramento della sicurezza del codice tramite il rilevamento e la correzione delle vulnerabilità di sicurezza.

Esempio del cliente

Nell'ambito della propria pipeline di integrazione continua, AnyCompany Retail effettua diversi tipi di test su tutti gli artefatti software. L'azienda lo sviluppo guidato dai test, per cui tutto il software è dotato di test di unità. Una volta costruito l'artefatto, eseguono dei test. end-to-end Al termine di questa prima serie di test, viene eseguita una scansione statica della sicurezza dell'applicazione, alla ricerca di vulnerabilità note. Gli sviluppatori ricevono messaggi al superamento di ciascun gate di test. Una volta completati tutti i test, l'artefatto software viene archiviato in un repository di artefatti.

Passaggi dell'implementazione

1. Collaborare con le parti interessate dell'organizzazione per sviluppare uno standard di test per gli artefatti software. Quali test standard devono superare tutti gli artefatti? Ci sono requisiti di conformità o di governance che devono essere inclusi nella copertura dei test? Devi condurre test di qualità del codice? Quando i test sono terminati, chi deve esserne a conoscenza?
    1. La AWS Deployment Pipeline Reference Architecture contiene un elenco autorevole dei tipi di test che possono essere condotti su artefatti software come parte di una pipeline di integrazione.
2. Dota la tua applicazione di strumenti con i test necessari in base allo standard di test del software. Ogni set di test deve essere completato in meno di dieci minuti. I test devono essere eseguiti come parte della pipeline di integrazione.
    a. Usa Amazon Q Developer, uno strumento di IA generativa utile per creare casi di test di unità (comprese le condizioni limite), generare funzioni utilizzando codice e commenti e implementare algoritmi noti.
    b. Usa Amazon CodeGuru Reviewer per testare il codice dell'applicazione alla ricerca di eventuali difetti.
    c. Puoi usare per AWS CodeBuild condurre i test su artefatti software.
    d. AWS CodePipeline può orchestrare i test software in una pipeline.

Risorse

Best practice correlate:

- OPS05-BP01 Usa il controllo della versione

- OPS05-BP06 Condividi gli standard di progettazione
- OPS05-BP07 Implementare pratiche per migliorare la qualità del codice
- OPS05- BP1 0 Integrazione e implementazione completamente automatizzate

Documenti correlati:

- Adozione di un approccio di sviluppo basato su test
- Accelerate your Software Development Lifecycle with Amazon Q
- Amazon Q Developer, now generally available, includes previews of new capabilities to reimagine developer experience
- Il cheat sheet definitivo per utilizzare Amazon Q Developer nel tuo IDE
- Shift-Left Workload, leveraging AI for Test Creation
- Amazon Q Developer Center
- 10 modi per creare applicazioni più velocemente con Amazon CodeWhisperer
- Oltre la copertura del codice con Amazon CodeWhisperer
- Le migliori pratiche per la progettazione tempestiva con Amazon CodeWhisperer
- Pipeline AWS CloudFormation di test automatizzata con e TaskCat CodePipeline
- Creazione di una pipeline end-to-end AWS DevSecOps CI/CD con strumenti e software open source SCA SAST DAST
- Getting started with testing serverless applications
- My CI/CD pipeline is my release captain
- Practicing Continuous Integration and Continuous Delivery on AWS Whitepaper

Video correlati:

- Implementa un agente di sviluppo software API con Amazon Q Developer Agent
- Installazione, configurazione e utilizzo di Amazon Q Developer con JetBrains IDEs (How-to)
- Padroneggiare l'arte di Amazon CodeWhisperer : playlist YouTube
- AWS re:Invent 2020: Infrastruttura testabile: test di integrazione su AWS
- AWS Summit ANZ 2021 - Promuovere una strategia orientata ai test e uno sviluppo basato sui test CDK
- Testa la tua infrastruttura come codice con AWS CDK

Risorse correlate:

- Creazione di applicazioni utilizzando l'intelligenza artificiale generativa con Amazon CodeWhisperer
- CodeWhisperer Workshop Amazon
- AWS Deployment Pipeline Reference Architecture - Application
- AWS Pipeline Kubernetes DevSecOps
- Workshop Policy come codice: sviluppo incentrato sui test
- Esegui test unitari per un'applicazione Node.js utilizzando GitHub AWS CodeBuild
- Use Serverspec for test-driven development of infrastructure code

Servizi correlati:

- Amazon Q Developer
- CodeGuru Revisore Amazon
- AWS CodeBuild
- AWS CodePipeline

OPS05-BP03 Utilizzare sistemi di gestione della configurazione

L'utilizzo di sistemi di gestione delle configurazioni permette di effettuare modifiche alle stesse e tenerne traccia. Questi sistemi riducono gli errori causati dai processi manuali e il livello di impegno richiesto per la distribuzione delle modifiche.

Durante l'inizializzazione di una risorsa, la gestione delle configurazioni statiche consente di impostare valori che dovrebbero rimanere coerenti per tutta la vita utile della risorsa. Al momento dell'inizializzazione, la gestione delle configurazioni dinamiche consente di impostare valori che possono cambiare nel corso della vita utile di una risorsa. Ad esempio, è possibile impostare un interruttore per la funzionalità in modo da attivarla nel codice tramite una modifica della configurazione o modificare il livello di dettaglio del log durante un incidente.

Le configurazioni vanno implementate in uno stato noto e coerente. Utilizza l'ispezione automatizzata per monitorare in modo continuo le configurazioni delle risorse tra ambienti e regioni. Occorre definire questi controlli come codice e gestione automatizzati per garantire l'applicazione coerente delle regole in tutti gli ambienti. Le modifiche alle configurazioni vanno aggiornate tramite procedure di controllo delle modifiche concordate e applicate in modo coerente, rispettando il controllo delle

versioni. Occorre gestire la configurazione dell'applicazione in modo indipendente rispetto al codice dell'applicazione e all'infrastruttura. In questo modo, si garantisce un'implementazione coerente tra più ambienti. Le modifiche alla configurazione non comportano la ricostruzione o la nuova implementazione dell'applicazione.

Risultato desiderato: puoi configurare, convalidare e implementare come parte della tua pipeline di integrazione continua e di distribuzione continua (CI/CD). Esegui il monitoraggio per verificare che le configurazioni siano corrette. Ciò riduce al minimo l'impatto sugli utenti finali e sui clienti.

Anti-pattern comuni:

- Aggiorni manualmente la configurazione del server Web all'interno del parco istanze e un certo numero di server non risponde a causa di errori di aggiornamento.
- Aggiorni manualmente il parco istanze del server applicazioni nel corso di molte ore. L'incoerenza nella configurazione durante la modifica causa comportamenti imprevisti.
- Qualcuno ha aggiornato i tuoi gruppi di sicurezza e i server Web non sono più accessibili. Senza sapere cosa è stato modificato, dedichi molto tempo a esaminare il problema prolungando il tempo necessario per il ripristino.
- Avvii una configurazione di preproduzione in produzione tramite CI/CD senza una convalida. Esponi utenti e clienti a dati e servizi errati.

Vantaggi dell'adozione di questa best practice: l'adozione di sistemi di gestione della configurazione riduce il livello di impegno necessario per apportare e tenere traccia delle modifiche e la frequenza degli errori causati dalle procedure manuali. I sistemi di gestione della configurazione forniscono garanzie per quanto riguarda la governance, la conformità e i requisiti normativi.

Livello di rischio associato se questa best practice non fosse adottata: medio

Guida all'implementazione

I sistemi di gestione della configurazione vengono utilizzati per tenere traccia e implementare le modifiche nelle configurazioni delle applicazioni e degli ambienti. I sistemi di gestione della configurazione vengono utilizzati anche per ridurre gli errori causati dai processi manuali, rendere le modifiche alla configurazione ripetibili e verificabili e per ridurre il livello di impegno.

Sì AWS, puoi utilizzarlo per monitorare continuamente AWS Configle configurazioni AWS delle risorse tra account e regioni. Questa soluzione aiuta a tenere traccia della cronologia delle configurazioni, a capire che effetto avrebbe la modifica di una configurazione sulle altre risorse e a

verificarle rispetto alle configurazioni previste o desiderate tramite Regole di AWS Config e pacchetti di conformitàAWS Config.

Per le configurazioni dinamiche delle tue applicazioni in esecuzione su EC2 istanze Amazon, contenitori AWS Lambda, applicazioni mobili o dispositivi IoT, puoi utilizzarle AWS AppConfigper configurarle, convalidarle, distribuirle e monitorarle nei tuoi ambienti.

Passaggi dell'implementazione

1. Identifica i proprietari della configurazione.
   a. Metti a conoscenza i proprietari delle configurazioni di qualsiasi esigenza di conformità, governance o normativa.
2. Identifica gli elementi e i risultati della configurazione.
   a. Gli elementi di configurazione sono tutte le configurazioni ambientali e dell'applicazione interessate da un'implementazione all'interno della pipeline CI/CD.
   b. I risultati finali includono criteri di successo, convalide e aspetti da monitorare.
3. Seleziona gli strumenti per la gestione della configurazione in base ai requisiti aziendali e alla pipeline di distribuzione.
4. Per modifiche significative alla configurazione, prendi in considerazione le implementazioni ponderate, ad esempio le distribuzioni canary, per ridurre al minimo l'impatto di configurazioni errate.
5. Integra la gestione della configurazione nella tua pipeline CI/CD.
6. Convalida tutte le modifiche inserite.

Risorse

Best practice correlate:

- OPS06-BP01 Piano per modifiche non riuscite
- OPS06-BP02 Implementazioni di test
- OPS06-BP03 Utilizzare strategie di implementazione sicure
- OPS06-BP04 Automatizza i test e il rollback

Documenti correlati:

- AWS Control Tower

- AWS Landing Zone Accelerator
- AWS Config
- Che cos'è AWS Config?
- AWS AppConfig
- Che cos'è AWS CloudFormation?
- Strumenti per sviluppatori in AWS

Video correlati:

- AWS re:Invent 2022 - Governance e conformità proattive per i carichi di lavoro AWS
- AWS re:Invent 2020: raggiungi la conformità come codice utilizzando il codice AWS Config
- Gestisci e distribuisci le configurazioni delle applicazioni con AWS AppConfig

OPS05-BP04 Usa sistemi di gestione della compilazione e dell'implementazione

Utilizza sistemi di gestione della creazione e implementazione. Questi sistemi riducono gli errori causati dai processi manuali e il livello di impegno richiesto per la distribuzione delle modifiche.

Nel AWS, puoi creare pipeline di integrazione continua/distribuzione continua (CI/CD) utilizzando servizi come AWS Developer Tools (ad esempio,,, e). AWS CodeCommitAWS CodeBuildAWS CodePipelineAWS CodeDeployAWS CodeStar

Risultato desiderato: i sistemi di gestione della costruzione e dell'implementazione supportano il sistema di distribuzione e integrazione continua (CI/CD) dell'organizzazione, che fornisce funzionalità per automatizzare rollout sicuri con le configurazioni corrette.

Anti-pattern comuni:

- Dopo aver compilato il codice nel sistema di sviluppo, copi il file eseguibile nei sistemi di produzione e questo non si avvia. I file di log locali indicano che l'operazione è risultata impossibile a causa della mancanza di dipendenze.
- Hai creato l'applicazione con nuove funzionalità nel tuo ambiente di sviluppo e fornisci il codice per eseguire il controllo qualità (QA). Il controllo qualità non riesce perché mancano asset statici.
- Venerdì, dopo un notevole sforzo, hai creato l'applicazione manualmente nel tuo ambiente di sviluppo, incluse le nuove funzionalità codificate. Lunedì non sei in grado di ripetere le fasi che ti hanno consentito di creare correttamente la tua applicazione.

- Esegui i test creati per la nuova versione. Quindi passi la settimana successiva a configurare un ambiente di test ed eseguire tutti i test di integrazione esistenti seguiti dai test delle prestazioni. Il nuovo codice ha un impatto inaccettabile sulle prestazioni e deve essere risviluppato e quindi ritestato.

Vantaggi dell'adozione di questa best practice fornendo meccanismi per gestire le attività di compilazione e implementazione, riduci il livello di impegno necessario per eseguire attività ripetitive, consenti ai membri del team di concentrarsi liberamente sulle loro attività creative di valore elevato e limiti l'introduzione di errori derivanti da procedure manuali.

Livello di rischio associato se questa best practice non fosse adottata: medio

Guida all'implementazione

I sistemi di gestione della creazione e implementazione vengono utilizzati per tenere traccia e implementare le modifiche, ridurre gli errori causati dai processi manuali e diminuire il livello di impegno richiesto per le implementazioni sicure. Automatizza completamente la pipeline di integrazione e implementazione dal check-in del codice fino alle fasi di creazione, test, implementazione e convalida. Ciò riduce il lead time e i costi, incoraggia una maggiore frequenza delle modifiche, riduce il livello di impegno e aumenta la collaborazione.

Passaggi dell'implementazione

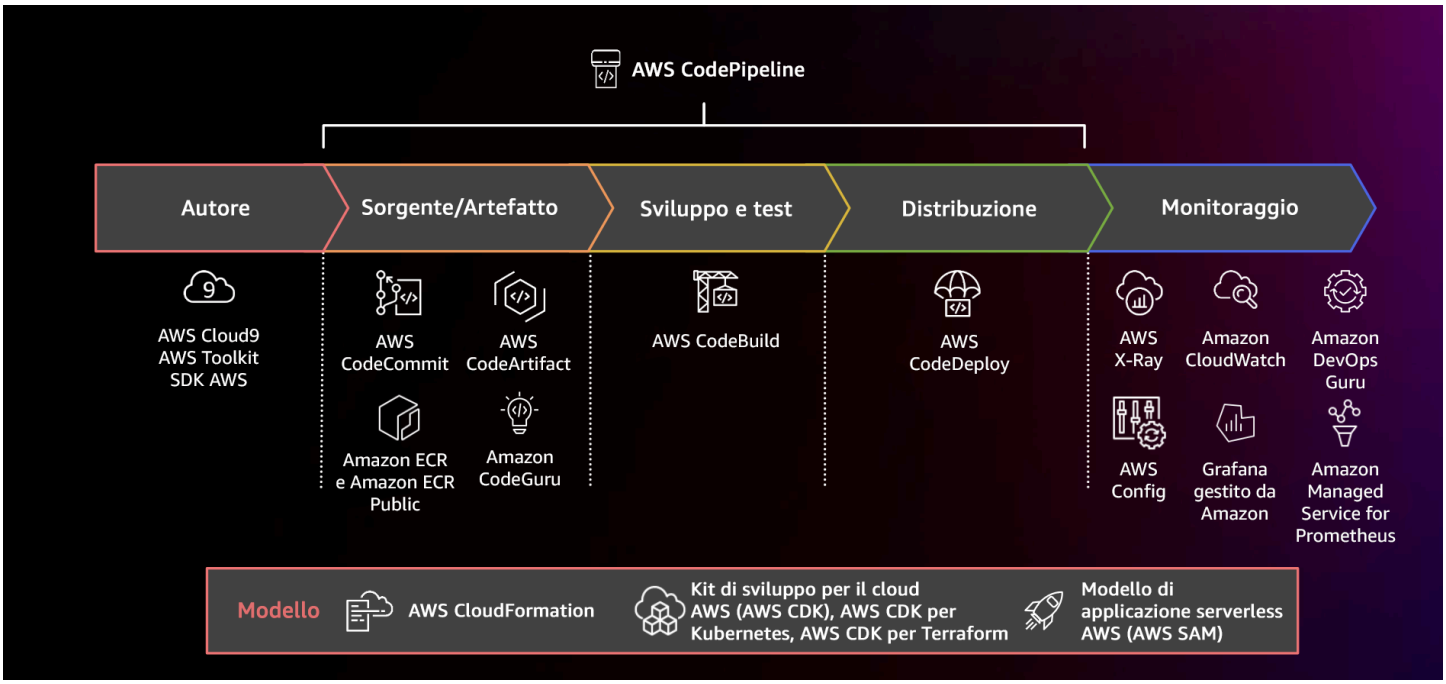


Diagramma che mostra una pipeline CI/CD che utilizza i servizi correlati AWS CodePipeline

1. AWS CodeCommit Utilizzatelo per controllare la versione, archiviare e gestire le risorse (come documenti, codice sorgente e file binari).
2. CodeBuild Utilizzatelo per compilare il codice sorgente, eseguire test unitari e produrre artefatti pronti per la distribuzione.
3. CodeDeploy Utilizzalo come servizio di distribuzione che automatizza le distribuzioni di applicazioni su istanze Amazon, EC2 istanze locali, funzioni serverless o Amazon. AWS Lambda ECS
4. Monitora le tue implementazioni.

Risorse

Best practice correlate:

- OPS06-BP04 Automatizza i test e il rollback

Documenti correlati:

- Strumenti per sviluppatori in AWS
- AWS CodeCommit Che cos'è?
- Che cos'è AWS CodeBuild?
- AWS CodeBuild
- Che cos'è AWS CodeDeploy?

Video correlati:

- AWS re:Invent 2022 - Le migliori pratiche di Well-Architected AWS per DevOps AWS

OPS05-BP05 Eseguire la gestione delle patch

La gestione delle patch consente di ottenere funzionalità, risolvere problemi e rispettare i requisiti di governance. Automatizza la gestione delle patch per ridurre gli errori causati dai processi manuali, dimensionare e ridurre il livello di impegno richiesto per applicare le patch.

La gestione delle patch e delle vulnerabilità fa parte delle attività di gestione dei rischi e dei vantaggi. È preferibile disporre di infrastrutture immutabili e distribuire carichi di lavoro in stati noti verificati. Se ciò non è realizzabile, l'applicazione di patch sul posto è l'alternativa.

Amazon EC2 Image Builder fornisce pipeline per aggiornare le immagini delle macchine. Come parte della gestione delle patch, considera Amazon Machine Images (AMIs) che utilizza una pipeline di AMI immagini o immagini di container con una pipeline di immagini Docker, AWS Lambda fornendo al contempo modelli per runtime personalizzati e librerie aggiuntive per rimuovere le vulnerabilità.

È necessario gestire gli aggiornamenti delle immagini di Amazon Machine Images per Linux o Windows Server utilizzando Amazon EC2 Image Builder. Puoi utilizzare Amazon Elastic Container Registry (AmazonECR) con la tua pipeline esistente per gestire ECS le immagini Amazon e gestire le EKS immagini Amazon. Lamba offre funzionalità di gestione delle versioni.

L'applicazione di patch non deve essere eseguita sui sistemi di produzione senza prima eseguire test in un ambiente sicuro. Le patch devono essere applicate solo se supportano risultati operativi o aziendali. È possibile utilizzare AWS Systems Manager Patch Manager per automatizzare il processo di applicazione delle patch ai sistemi gestiti e pianificare l'attività utilizzando Systems Manager Maintenance Windows. AWS

Risultato desiderato: le tue immagini AMI e quelle del contenitore sono state patchate e pronte per il lancio. up-to-date È possibile tenere traccia dello stato di tutte le immagini implementate e conoscere la conformità delle patch. Puoi eseguire report sullo stato attuale e disporre di un processo per soddisfare le tue esigenze di conformità.

Anti-pattern comuni:

- Ti viene assegnato il compito di applicare tutte le nuove patch di sicurezza entro 2 ore, il che provoca più interruzioni a causa dell'incompatibilità dell'applicazione con le patch.
- Una libreria senza patch comporta conseguenze indesiderate in quanto parti sconosciute utilizzano vulnerabilità al suo interno per accedere al carico di lavoro.
- L'applicazione di patch agli ambienti per sviluppatori viene eseguita automaticamente senza avvisare gli sviluppatori. Gli sviluppatori ti inviano più reclami perché il loro ambiente non funziona come previsto.
- Non hai applicato una patch al off-the-shelf software commerciale su un'istanza persistente. Quando hai problemi con il software e contatti il fornitore, questo ti informa che la versione non è supportata e che devi applicare le patch a un livello specifico per ricevere assistenza.

- Una patch rilasciata di recente per il software di crittografia utilizzato offre miglioramenti significativi in termini di prestazioni. Il sistema privo di patch presenta problemi di prestazioni che rimangono in vigore a causa della mancata applicazione di patch.
- Ricevi una notifica di una vulnerabilità zero-day che richiede una correzione di emergenza; quindi devi applicare manualmente le patch a tutti i tuoi ambienti.

Vantaggi dell'adozione di questa best practice: stabilendo un processo di gestione delle patch, inclusi i criteri per l'applicazione di patch e la metodologia di distribuzione tra gli ambienti, sarai in grado di dimensionare e generare report sui livelli di patch. Ciò fornisce garanzie sull'applicazione delle patch di sicurezza e una chiara visibilità sullo stato delle correzioni note in atto. Ciò incoraggia l'adozione delle caratteristiche e funzionalità desiderate, aiuta a eliminare rapidamente i problemi e a mantenere la conformità alla governance. Implementa sistemi di gestione delle patch e automazione per ridurre il livello di impegno per distribuire le patch e limitare gli errori causati dai processi manuali.

Livello di rischio associato se questa best practice non fosse adottata: medio

Guida all'implementazione

Applica patch ai sistemi per correggere gli errori, ottenere le funzionalità o le capacità desiderate e assicurare la conformità alle policy di governance e ai requisiti di supporto del fornitore. Nei sistemi immutabili, distribuisci con il set di patch appropriato per raggiungere il risultato desiderato. Automatizza il meccanismo di gestione delle patch per ridurre il tempo necessario per applicare le patch, evitare gli errori causati dai processi manuali e diminuire il livello di impegno richiesto per applicare le patch.

Passaggi dell'implementazione

Per Amazon EC2 Image Builder:

1. Utilizzando Amazon EC2 Image Builder, specifica i dettagli della pipeline:
    a. Crea una pipeline di immagini e assegnale un nome.
    b. Definisci la pianificazione e il fuso orario della pipeline.
    c. Configura eventuali dipendenze.
2. Scegli una ricetta:
    a. Seleziona una ricetta esistente o creane una nuova.
    b. Seleziona il tipo di immagine.
    c. Assegna un nome e una versione alla tua ricetta.

   d. Seleziona l'immagine di base.
   e. Aggiungi componenti di compilazione e inseriscili nel registro di destinazione.
3. Facoltativo: definisci la configurazione dell'infrastruttura.
4. Facoltativo: definisci le impostazioni di configurazione.
5. Verifica le impostazioni.
6. Mantieni il livello di igiene delle ricette a livelli ottimali.

Per Gestione patch di Systems Manager:

1. Crea una baseline delle patch.
2. Seleziona un metodo di applicazione delle patch.
3. Abilita il report e la scansione della conformità.

Risorse

Best practice correlate:

- OPS06-BP04 Automatizza i test e il rollback

Documenti correlati:

- Cos'è Amazon EC2 Image Builder
- Crea una pipeline di immagini utilizzando Amazon EC2 Image Builder
- Create a container image pipeline
- AWS Systems Manager Patch Manager
- Working with Patch Manager
- Working with patch compliance reports
- AWS Strumenti per sviluppatori

Video correlati:

- CI/CD per applicazioni serverless su AWS
- Design with Ops in Mind

Esempi correlati:

- Well-Architected Labs: inventario e gestione delle patch
- AWS Systems Manager Tutorial Patch Manager

OPS05-BP06 Condividi gli standard di progettazione

Condividi le best practice con i team per incrementare la consapevolezza e potenziare al massimo i vantaggi delle attività di sviluppo. Documentale e mantienile aggiornate di pari passo con l'evoluzione dell'architettura. Se nella tua organizzazione vengono applicati standard condivisi, è fondamentale che esistano meccanismi per richiedere aggiunte, modifiche ed eccezioni agli standard. Senza questa opzione, gli standard diventano un ostacolo per l'innovazione.

Risultato desiderato: gli standard di progettazione vengono condivisi fra team nelle organizzazioni. Sono documentati e conservati up-to-date man mano che le migliori pratiche si evolvono.

Anti-pattern comuni:

- Due team di sviluppo hanno creato ciascuno un servizio di autenticazione utente. Gli utenti devono mantenere un set separato di credenziali per ogni parte del sistema a cui vogliono accedere.
- Ogni team gestisce la propria infrastruttura. Un nuovo requisito di conformità impone una modifica all'infrastruttura e ogni team la implementa in modo diverso.

Vantaggi dell'adozione di questa best practice: l'uso di standard condivisi incoraggia l'applicazione di best practice e permette di ottenere i massimi vantaggi dalle attività di sviluppo. La documentazione e l'aggiornamento degli standard di progettazione consentono all'organizzazione di attenersi up-to-date alle migliori pratiche e ai requisiti di sicurezza e conformità.

Livello di rischio associato se questa best practice non fosse adottata: medio

Guida all'implementazione

Condividi le best practice, gli standard di progettazione, gli elenchi di controllo, le procedure operative, le linee guida e i requisiti di governance esistenti tra team diversi. Definisci procedure per richiedere modifiche, aggiunte ed eccezioni agli standard di progettazione per supportare il miglioramento e l'innovazione. Rendi noto ai team il contenuto pubblicato. Disponete di un meccanismo per mantenere gli standard di progettazione up-to-date man mano che emergono nuove best practice.

Esempio del cliente

AnyCompany Retail dispone di un team di architettura interfunzionale che crea modelli di architettura software. Questo team crea l'architettura con conformità e governance integrate. I team che adottano gli standard condivisi traggono vantaggio dall'integrazione di conformità e governance. Possono creare rapidamente soluzioni sulla base degli standard di progettazione. Il team responsabile dell'architettura si incontra ogni trimestre per valutare i modelli architetturali e aggiornarli, se necessario.

Passaggi dell'implementazione

1. Identifica un team interfunzionale responsabile dello sviluppo e dell'aggiornamento degli standard di progettazione. Questo team collaborerà con le parti interessate in tutta l'organizzazione per sviluppare standard di progettazione, procedure operative, elenchi di controllo, linee guida e requisiti di governance. Documenta gli standard di progettazione e condividili internamente all'organizzazione.
    a. AWS Service Catalog può aiutarti a creare portfolio che rappresentano gli standard di progettazione usando il modello Infrastructure as code (IaC). Puoi condividere portfolio tra più account.
2. Disponete di un meccanismo per mantenere gli standard di progettazione up-to-date man mano che vengono identificate nuove best practice.
3. Se gli standard di progettazione vengono applicati a livello centrale, definisci un processo per richiedere modifiche, aggiornamenti ed eccezioni.

Livello di impegno per il piano di implementazione: medio Lo sviluppo di un processo per creare e condividere standard di progettazione può richiedere il coordinamento e la cooperazione con le parti interessate in tutta l'organizzazione.

Risorse

Best practice correlate:

- OPS01-BP03 Valuta i requisiti di governance: i requisiti di governance influiscono sugli standard di progettazione.
- OPS01-BP04 Valuta i requisiti di conformità: la conformità è un fattore essenziale nella creazione di standard di progettazione.

- OPS07-BP02 Garantire una revisione coerente della prontezza operativa: gli elenchi di controllo della prontezza operativa sono un meccanismo per implementare standard di progettazione durante la progettazione del carico di lavoro.
- OPS11-BP01 Avere un processo per il miglioramento continuo: l'aggiornamento degli standard di progettazione contribuisce a un miglioramento continuo.
- OPS11-BP04 Eseguire la gestione della conoscenza: nell'ambito della procedura di gestione delle informazioni, documenta e condividi gli standard di progettazione.

Documenti correlati:

- AWS Backup Automatizzaci con AWS Service Catalog
- AWS Service Catalog Account Factory-Enhanced
- In che modo Expedia Group ha creato l'offerta Database as a Service () utilizzando DBaaS AWS Service Catalog
- Maintain visibility over the use of cloud architecture patterns
- Semplifica la condivisione dei tuoi AWS Service Catalog portafogli in un'unica configurazione AWS Organizations

Video correlati:

- AWS Service Catalog — Guida introduttiva
- AWS re:Invent 2020: gestisci i tuoi AWS Service Catalog portafogli come un esperto

Esempi correlati:

- AWS Service Catalog Architettura di riferimento
- AWS Service Catalog Workshop

Servizi correlati:

- AWS Service Catalog

OPS05-BP07 Implementare pratiche per migliorare la qualità del codice

Implementa prassi per migliorare la qualità del codice e ridurre al minimo i difetti. Alcuni esempi includono sviluppo basato su test, revisioni del codice, adozione degli standard e programmazione in coppia. Inserisci queste prassi nel processo di integrazione continua e distribuzione continua.

Risultato desiderato: la tua organizzazione usa best practice come le revisioni del codice e la programmazione in coppia per migliorare la qualità del codice. Sviluppatori e operatori adottano le best practice per la qualità del codice nell'ambito del ciclo di vita di sviluppo del software.

Anti-pattern comuni:

- Commit del codice nel ramo principale dell'applicazione senza alcuna revisione. In questo modo, la modifica viene implementata in automatico nell'ambiente di produzione e causa un'interruzione.
- Una nuova applicazione viene sviluppata senza test di unità o di integrazione end-to-end. Non è possibile in alcun modo testare l'applicazione prima dell'implementazione.
- I team apportano modifiche manuali nell'ambiente di produzione per gestire gli errori. Le modifiche non vengono sottoposte a test o revisioni del codice, né vengono acquisite o registrate durante i processi di integrazione continua e distribuzione continua.

Vantaggi dell'adozione di questa best practice: l'adozione di pratiche per migliorare la qualità del codice ti consente di ridurre al minimo i problemi di produzione. La qualità del codice semplifica l'uso delle best practice, come la programmazione in coppia, le revisioni del codice e l'implementazione di strumenti di produttività basati sull'IA.

Livello di rischio associato se questa best practice non fosse adottata: medio

Guida all'implementazione

Implementa prassi per migliorare la qualità del codice in modo da ridurre gli errori prima dell'implementazione. Usa prassi come lo sviluppo basato su test, le revisioni del codice e la programmazione in coppia per migliorare la qualità dello sviluppo.

Usa la potenza dell'IA generativa con Amazon Q Developer per migliorare la produttività degli sviluppatori e la qualità del codice. Amazon Q Developer include la generazione di suggerimenti di codice (basati su modelli linguistici di grandi dimensioni), la produzione di test di unità (comprese le condizioni limite) e il miglioramento della sicurezza del codice tramite il rilevamento e la correzione delle vulnerabilità di sicurezza.

Esempio del cliente

AnyCompany La vendita al dettaglio adotta diverse pratiche per migliorare la qualità del codice. L'azienda ha adottato lo sviluppo basato su test come standard per la scrittura di applicazioni. Per alcune nuove funzionalità, gli sviluppatori eseguiranno la programmazione in coppia durante uno sprint. Ogni richiesta pull viene sottoposta a una revisione del codice da parte di uno sviluppatore senior prima di essere integrata e implementata.

Passaggi dell'implementazione

1. Adotta prassi per la qualità del codice come lo sviluppo basato su test, le revisioni del codice e la programmazione in coppia nel processo di integrazione continua e distribuzione continua. Usa queste tecniche per migliorare la qualità del software.
   a. Usa Amazon Q Developer, uno strumento di IA generativa utile per creare casi di test di unità (comprese le condizioni limite), generare funzioni utilizzando codice e commenti, implementare algoritmi noti, rilevare violazioni e vulnerabilità delle policy di sicurezza nel codice, rilevare segreti, effettuare la scansione dell'infrastructure as code (IaC), documentare il codice e apprendere più rapidamente le librerie di codici di terze parti.
   b. Amazon CodeGuru Reviewer può fornire consigli di programmazione per il codice Java e Python utilizzando l'apprendimento automatico.
   c. Puoi creare ambienti di sviluppo condivisi con AWS Cloud9 in cui collaborare allo sviluppo del codice.

Livello di impegno per il piano di implementazione: medio Esistono molti modi per implementare questa best practice, ma la realizzazione dell'adozione da parte dell'organizzazione può essere problematica.

Risorse

Best practice correlate:

- OPS05-BP02 Verifica e convalida le modifiche
- OPS05-BP06 Condividi gli standard di progettazione

Documenti correlati:

- Adozione di un approccio di sviluppo basato su test
- Accelerate your Software Development Lifecycle with Amazon Q

- Amazon Q Developer, now generally available, includes previews of new capabilities to reimagine developer experience
- Il cheat sheet definitivo per utilizzare Amazon Q Developer nel tuo IDE
- Shift-Left Workload, leveraging AI for Test Creation
- Amazon Q Developer Center
- 10 modi per creare applicazioni più velocemente con Amazon CodeWhisperer
- Oltre la copertura del codice con Amazon CodeWhisperer
- Le migliori pratiche per la progettazione tempestiva con Amazon CodeWhisperer
- Agile Software Guide
- My CI/CD pipeline is my release captain
- Automatizza le revisioni del codice con Amazon Reviewer CodeGuru
- Adozione di un approccio di sviluppo basato su test
- Come DevFactory creare applicazioni migliori con Amazon CodeGuru
- On Pair Programming
- RENGAInc. automatizza le revisioni dei codici con Amazon CodeGuru
- The Art of Agile Development: Test-Driven Development
- Why code reviews matter (and actually save time!)

Video correlati:

- Implementa un agente di sviluppo software API con Amazon Q Developer Agent
- Installazione, configurazione e utilizzo di Amazon Q Developer con JetBrains IDEs (How-to)
- Padroneggiare l'arte di Amazon CodeWhisperer : playlist YouTube
- AWS re:Invent 2020: miglioramento continuo della qualità del codice con Amazon CodeGuru
- AWS Summit ANZ 2021 - Promuovere una strategia basata sui test e uno sviluppo basato sui test CDK

Servizi correlati:

- Amazon Q Developer
- CodeGuru Revisore Amazon
- Amazon CodeGuru Profiler

- AWS Cloud9

OPS05-BP08 Usa più ambienti

Utilizza più ambienti per sperimentare, sviluppare e testare il carico di lavoro. Applica livelli crescenti di controlli man mano che gli ambienti si avvicinano alla fase di produzione per avere la certezza che il carico di lavoro funzioni come previsto una volta implementato.

Risultato desiderato: disponi di più ambienti che riflettono le tue esigenze di conformità e governance. Testi e promuovi il codice negli ambienti lungo il tuo percorso verso la produzione.

Anti-pattern comuni:

- Stai sviluppando in un ambiente di sviluppo condiviso e un altro sviluppatore sovrascrive le tue modifiche al codice.
- I controlli di sicurezza restrittivi nell'ambiente di sviluppo condiviso impediscono di sperimentare nuovi servizi e funzionalità.
- Esegui test di carico sui tuoi sistemi di produzione e causa un'interruzione per i tuoi utenti.
- Si è verificato un errore critico che ha causato la perdita di dati nella produzione. Nel tuo ambiente di produzione tenti di ricreare le condizioni che portano alla perdita di dati in modo da poter identificare come si è verificata e impedire che si ripeta. Per evitare un'ulteriore perdita di dati durante il test, devi rendere l'applicazione non disponibile per i tuoi utenti.
- Stai operando un servizio multi-tenant e non sei in grado di supportare la richiesta di un cliente per un ambiente dedicato.
- Ogni volta che esegui un test, lo fai nel tuo ambiente di produzione.
- Ritieni che la semplicità di un singolo ambiente prevalga sulla portata dell'impatto che possono avere modifiche all'interno dell'ambiente.

Vantaggi dell'adozione di questa best practice: puoi supportare più ambienti di sviluppo, test e produzione simultanei senza creare conflitti tra sviluppatori o community di utenti.

Livello di rischio associato se questa best practice non fosse adottata: medio

Guida all'implementazione

Utilizza più ambienti e fornisci agli sviluppatori ambienti sandbox con controlli minimi per incoraggiare la sperimentazione. Fornisci ambienti di sviluppo individuali per facilitare il lavoro in parallelo, incrementando l'agilità dello sviluppo. Implementa controlli più rigorosi negli ambienti che si

avvicinano alla produzione per consentire agli sviluppatori di innovare. Utilizza l'approccio Infrastructure as code e sistemi di gestione delle configurazioni per distribuire ambienti configurati in modo coerente con i controlli presenti in produzione per assicurare che i sistemi funzionino nel modo previsto quando vengono distribuiti. Quando gli ambienti non vengono utilizzati, disattivali per evitare costi associati alle risorse inattive, ad esempio i sistemi di sviluppo nelle ore serali e nei fine settimana. Durante i test di carico, è necessario implementare ambienti equivalenti a quelli di produzione per migliorare la validità dei risultati.

Risorse

Documenti correlati:

- Instance Scheduler attivo AWS
- Che cos'è AWS CloudFormation?

OPS05-BP09 Apporta modifiche frequenti, piccole e reversibili

Le modifiche frequenti, minime e reversibili riducono la portata e l'impatto di una modifica. Le modifiche frequenti, minime e reversibili, se effettuate utilizzando congiuntamente sistemi di gestione delle modifiche, di gestione della configurazione e di compilazione e distribuzione, riducono la portata e l'impatto di una modifica. Questo si traduce in una risoluzione dei problemi più efficace, accelerando la correzione e mantenendo la possibilità di rollback delle modifiche.

Anti-pattern comuni:

- Distribuisci una nuova versione della tua applicazione ogni trimestre con una finestra di modifica, il che comporta la disattivazione di un servizio di base.
- Spesso apporti modifiche allo schema del database senza che ne venga tenuta traccia nei sistemi di gestione.
- Esegui aggiornamenti manuali sul posto, sovrascrivendo le installazioni e le configurazioni esistenti, senza avere un chiaro piano di rollback.

Vantaggi dell'adozione di questa best practice: velocizzazione degli sforzi di sviluppo grazie all'implementazione frequente di piccole modifiche. Quando le modifiche sono minime, è molto più semplice identificare se hanno conseguenze indesiderate e, in tal caso, ripristinare la condizione precedente. Quando le modifiche sono reversibili, il rischio di implementare le modifiche è minore in quanto il ripristino è semplificato. Il processo di modifica comporta un rischio ridotto e l'impatto di una modifica non corretta è ridotto.

Livello di rischio associato se questa best practice non fosse adottata: basso

Guida all'implementazione

Applica modifiche frequenti, minime e reversibili per ridurre la portata e l'impatto di una modifica. In questo modo si semplifica la risoluzione dei problemi, si velocizza la correzione ed è possibile eseguire il rollback di una modifica. Inoltre, aggiunge più rapidamente valore al business.

Risorse

Best practice correlate:

- OPS05-BP03 Utilizzare sistemi di gestione della configurazione
- OPS05-BP04 Usa sistemi di gestione della compilazione e dell'implementazione
- OPS06-BP04 Automatizza i test e il rollback

Documenti correlati:

- Implementazione di microservizi su AWS
- Microservices - Observability

OPS05- BP1 0 Integrazione e implementazione completamente automatizzate

Automatizza la creazione, l'implementazione e il test del carico di lavoro. Questo riduce gli errori causati dai processi manuali e l'impegno necessario per distribuire le modifiche.

Applica i metadati utilizzando i tag delle risorse e gli AWS Resource Groups seguendo una strategia di applicazione dei tag coerente per consentire l'identificazione delle risorse. Applica tag alle risorse per organizzare, monitorare i costi e controllare gli accessi e ottimizza l'esecuzione delle attività operative automatizzate.

Risultato desiderato: chi si occupa di sviluppo utilizza strumenti per distribuire codice ed effettuare la promozione a produzione. Gli sviluppatori non devono accedere a per fornire AWS Management Console gli aggiornamenti. Esiste un audit trail completo di modifiche e configurazioni che soddisfa le esigenze di governance e conformità. I processi sono ripetibili e standardizzati tra i team. Gli sviluppatori sono liberi di concentrarsi sullo sviluppo e sui rilasci del codice, aumentando la produttività.

Anti-pattern comuni:

- Venerdì termini la creazione del nuovo codice per il ramo delle funzionalità. Lunedì, dopo aver eseguito gli script di test di qualità del codice e tutti gli script dei test di unità, effettui il check-in del codice per il prossimo rilascio programmato.
- Ti verrà assegnato di codificare una correzione per un problema critico che interessa un numero elevato di clienti nella produzione. Dopo aver testato la correzione, esegui il commit del codice e richiedi via e-mail alla gestione delle modifiche l'approvazione per implementarlo in produzione.
- In qualità di sviluppatore, accedi AWS Management Console a per creare un nuovo ambiente di sviluppo utilizzando metodi e sistemi non standard.

Vantaggi dell'adozione di questa best practice: implementando sistemi di gestione automatizzati di compilazione e implementazione, si riduce il numero di errori causati dai processi manuali e lo sforzo di implementare le modifiche aiutando i membri del team a concentrarsi sull'offerta di valore aggiunto. Maggiore velocità di consegna man mano che procedi verso la promozione a produzione.

Livello di rischio associato se questa best practice non fosse adottata: basso

Guida all'implementazione

Utilizza i sistemi di gestione della compilazione e implementazione per tenere traccia e implementare le modifiche, ridurre gli errori causati dai processi manuali e ridurre il livello di impegno richiesto. Automatizza completamente la pipeline di integrazione e implementazione dal check-in del codice fino alle fasi di creazione, test, implementazione e convalida. In questo modo è possibile diminuire il lead time, incoraggiare una maggiore frequenza di modifica, ridurre il livello di impegno e accelerare il time-to-market, il che si traduce in una maggiore produttività e in un aumento della sicurezza del codice man mano che procedi con la promozione verso la produzione.

Risorse

Best practice correlate:

- OPS05-BP03 Utilizzare sistemi di gestione della configurazione
- OPS05-BP04 Usa sistemi di gestione della compilazione e dell'implementazione

Documenti correlati:

- Che cos'è AWS CodeBuild?
- Che cos'è AWS CodeDeploy?

Video correlati:

- AWS re\ :Invent 2022 - Le migliori pratiche di AWS Well-Architected per DevOps AWS

OPS6. In che modo mitighi i rischi dell'implementazione?

Adotta approcci per fornire un feedback rapido sulla qualità e che permettano un ripristino veloce dalle modifiche che non hanno i risultati previsti. L'uso di queste prassi consente di mitigare l'impatto dei problemi introdotti attraverso l'implementazione delle modifiche.

Best practice

- OPS06-BP01 Piano per modifiche non riuscite
- OPS06-BP02 Implementazioni di test
- OPS06-BP03 Utilizzare strategie di implementazione sicure
- OPS06-BP04 Automatizza i test e il rollback

OPS06-BP01 Piano per modifiche non riuscite

Pianifica il ripristino di uno stato corretto noto o la correzione nell'ambiente di produzione nel caso in cui l'implementazione generi un risultato indesiderato. Disporre di una policy per stabilire un piano di questo tipo aiuta tutti i team a sviluppare strategie di ripristino dalle modifiche con esito negativo. Alcune strategie di esempio sono le fasi di implementazione e rollback, le policy di modifica, i flag di funzionalità, l'isolamento del traffico e lo spostamento del traffico. Una singola release può includere più modifiche ai componenti correlati. La strategia dovrebbe fornire la capacità di resistere o ripristinare in caso di guasto generato da qualsiasi modifica dei componenti.

Risultato desiderato: hai preparato un piano di ripristino dettagliato per la modifica in caso di fallimento. Inoltre, hai ridotto le dimensioni della release per ridurre al minimo il potenziale impatto su altri componenti del carico di lavoro. Di conseguenza, hai ridotto l'impatto aziendale abbreviando i potenziali tempi di inattività causati da una modifica non riuscita e aumentando la flessibilità e l'efficienza dei tempi di ripristino.

Anti-pattern comuni:

- Hai eseguito un'implementazione e l'applicazione è diventata instabile, ma sembra che ci siano utenti attivi sul sistema. Devi decidere se eseguire il rollback della modifica e influire sugli utenti attivi o aspettare di eseguire il rollback della modifica, sapendo che gli utenti potranno essere comunque influenzati.

- Dopo aver apportato una modifica di routine, i nuovi ambienti sono accessibili, ma una delle sottoreti è diventata irraggiungibile. Devi decidere se eseguire il rollback di tutto o provare a correggere il problema della sottorete inaccessibile. Mentre prendi tale decisione, la sottorete rimane irraggiungibile.
- I tuoi sistemi non sono progettati in modo da consentire loro di essere aggiornati con release più piccole. Di conseguenza, è difficile annullare tali modifiche di massa (bulk changes) durante un'implementazione conclusasi con esito negativo.
- Non utilizzi il modello Infrastructure as code (IaC) e hai apportato aggiornamenti manuali all'infrastruttura che hanno portato a configurazioni indesiderate. Non è possibile tracciare e ripristinare in modo efficace le modifiche manuali.
- Poiché non hai misurato l'aumento della frequenza delle implementazioni, il tuo team non è incentivato a ridurre le dimensioni delle modifiche e a migliorare i piani di rollback per ogni modifica, con conseguente aumento dei rischi e dei tassi di fallimento.
- Non misuri la durata totale di un'interruzione causata da modifiche con esito negativo. Il tuo team non è in grado di stabilire le priorità e migliorare il processo di implementazione e l'efficacia del piano di ripristino.

Vantaggi derivanti dall'adozione di questa procedura ottimale: disporre di un piano di ripristino in caso di modifiche non riuscite riduce al minimo il tempo medio di ripristino () MTTR e riduce l'impatto aziendale.

Livello di rischio associato se questa best practice non fosse adottata: elevato

Guida all'implementazione

Una policy e una pratica coerenti e documentate adottate dai team di rilascio consentono a un'organizzazione di pianificare cosa dovrebbe succedere in caso di modifiche con esito negativo. In circostanze specifiche la policy dovrebbe consentire la possibilità di apportare correzioni per garantire la prosecuzione del processo. In entrambe le situazioni, un piano di correzione (fix forward) o ripristino (rollback) deve essere ben documentato e testato prima dell'implementazione nei sistemi di produzione live, in modo da ridurre al minimo il tempo necessario per ripristinare una modifica.

Passaggi dell'implementazione

1. Documenta le policy che richiedono ai team di disporre di piani efficaci per invertire le modifiche entro un periodo di tempo specificato.

   a. Le policy devono specificare quando è consentita una situazione di applicazione di correzioni per garantire la prosecuzione del processo.
   b. Richiedi un piano di rollback documentato che sia accessibile a tutti i soggetti coinvolti.
   c. Specifica i requisiti per il rollback (ad esempio, quando si rileva che sono state implementate modifiche non autorizzate).
2. Analizza il livello di impatto di tutte le modifiche relative a ciascun componente di un carico di lavoro.
   a. Consenti che le modifiche ripetibili siano standardizzate, basate su modelli e preautorizzate se seguono un flusso di lavoro coerente che applica le policy di modifica.
   b. Riduci il potenziale impatto di qualsiasi modifica riducendone le dimensioni, in modo che il ripristino richieda meno tempo e abbia un impatto aziendale minore.
   c. Assicurati che le procedure di rollback riportino il codice allo stato corretto noto per evitare incidenti, ove possibile.
3. Integra strumenti e flussi di lavoro per applicare le tue policy in modo programmatico.
4. Rendi visibili i dati sulle modifiche agli altri responsabili di carichi di lavoro per migliorare la velocità di diagnosi di eventuali modifiche con esito negativo che non possono essere ripristinate.
   a. Misura il successo di questa pratica utilizzando dati di modifica visibili e identifica miglioramenti iterativi.
5. Utilizza gli strumenti di monitoraggio per verificare il successo o il fallimento di un'implementazione per accelerare il processo decisionale sul rollback.
6. Misura la durata dell'interruzione durante una modifica con esito negativo per migliorare continuamente i tuoi piani di ripristino.

Livello di impegno per il piano di implementazione: medio

Risorse

Best practice correlate:

- OPS06-BP04 Automatizza i test e il rollback

Documenti correlati:

- AWS Builders Library | Garantire la sicurezza del rollback durante le implementazioni
- AWS Whitepaper | Gestione delle modifiche nel cloud

Video correlati:

- re:Invent 2019 | Amazon's approach to high-availability deployment

OPS06-BP02 Implementazioni di test

Testa le procedure di rilascio in pre-produzione utilizzando la stessa configurazione di implementazione, i controlli di sicurezza, i passaggi e le procedure utilizzati nell'ambiente di produzione. Verifica che tutte le fasi implementate siano state completate come previsto, ad esempio l'ispezione di file, configurazioni e servizi. Verifica ulteriormente tutte le modifiche con test funzionali, di integrazione e di carico, oltre ad attivare tutte le attività di monitoraggio come i controlli dell'integrità. Eseguendo questi test, è possibile identificare tempestivamente i problemi di implementazione con l'opportunità di pianificarli e mitigarli prima del passaggio nell'ambiente di produzione.

Puoi creare ambienti paralleli temporanei per testare ogni modifica. Automatizza l'implementazione degli ambienti di test utilizzando il modello Infrastructure as code (IaC) per ridurre la quantità di lavoro necessaria e garantire stabilità, coerenza e una distribuzione più rapida delle funzionalità.

Risultato desiderato: la tua organizzazione adotta una cultura di sviluppo che include il test delle implementazioni. Ciò garantisce che i team siano concentrati sulla realizzazione di valore aziendale anziché sulla gestione delle release. I team vengono coinvolti fin dall'identificazione dei rischi di implementazione per determinare il percorso di mitigazione appropriato.

Anti-pattern comuni:

- Durante le release di produzione, le implementazioni non testate causano problemi frequenti che richiedono una risoluzione mirata e l'escalation.
- La tua release contiene porzioni del modello Infrastructure as code (IaC) che aggiornano le risorse esistenti. Non sei sicuro che l'IaC funzionerà correttamente e non avrà un impatto sulle risorse.
- Viene implementata una nuova funzionalità interessante nella tua applicazione. Non funziona come previsto e non c'è visibilità finché non viene segnalata dagli utenti interessati.
- I certificati vengono aggiornati. Si installano accidentalmente i certificati sui componenti sbagliati, il che non viene rilevato e influisce sui visitatori poiché non è possibile stabilire una connessione sicura al sito web.

Vantaggi dell'adozione di questa best practice: test approfonditi in fase di pre-produzione delle procedure di implementazione e delle modifiche da queste introdotte riducono al minimo il potenziale

impatto sulla produzione causato dalle fasi di implementazione. Ciò aumenta la fiducia durante il rilascio in produzione e riduce al minimo la necessità di supporto operativo senza rallentare la velocità di distribuzione delle modifiche apportate.

Livello di rischio associato se questa best practice non fosse adottata: elevato

Guida all'implementazione

Testare il processo di implementazione è importante quanto testare le modifiche derivanti dall'implementazione. Ciò può essere ottenuto testando le fasi di implementazione in un ambiente di pre-produzione che rispecchi il più fedelmente possibile quello di produzione. I problemi più comuni, come fasi di implementazione incomplete o contenenti errori o configurazioni errate, possono essere individuati di conseguenza prima di passare all'ambiente di produzione. Inoltre, è possibile testare le fasi di ripristino.

Esempio del cliente

Nell'ambito della propria pipeline di integrazione e distribuzione continua (CI/CD), AnyCompany Retail esegue le fasi definite necessarie per rilasciare aggiornamenti dell'infrastruttura e del software per i propri clienti in un ambiente simile a quello di produzione. La pipeline comprende controlli preliminari per rilevare le deviazioni (il rilevamento delle modifiche alle risorse eseguite al di fuori dell'IaC) nelle risorse prima dell'implementazione, nonché per convalidare le azioni che l'IaC intraprende al suo avvio. Convalida le fasi dell'implementazione, ad esempio la verifica che determinati file e configurazioni siano presenti e che i servizi siano in esecuzione e rispondano correttamente ai controlli dell'integrità sull'host locale, prima di effettuare nuovamente la registrazione sul bilanciatore del carico. Inoltre, tutte le modifiche attivano una serie di test automatici, come test funzionali, di sicurezza, di regressione, di integrazione e di carico.

Passaggi dell'implementazione

1. Esegui controlli di pre-installazione per rispecchiare l'ambiente di pre-produzione in produzione.
   a. Utilizza il rilevamento della deriva per rilevare quando le risorse sono state modificate all'esterno. AWS CloudFormation
   b. Utilizzate i set di modifiche per verificare che l'intento di un aggiornamento dello stack corrisponda alle azioni AWS CloudFormation intraprese all'avvio del set di modifiche.
2. Ciò attiva una fase di approvazione manuale in AWS CodePipeline per autorizzare l'implementazione nell'ambiente di preproduzione.
3. Utilizza configurazioni di distribuzione, come AWS CodeDeploy AppSpeci file, per definire le fasi di distribuzione e convalida.

4. Ove applicabile, esegui AWS CodeDeploy l'integrazione con altri AWS servizi o AWS CodeDeploy con prodotti e servizi dei partner.
5. Monitora le distribuzioni utilizzando Amazon e CloudWatch le AWS CloudTrail notifiche SNS degli eventi di Amazon.
6. Esegui test automatici post-implementazione, inclusi test funzionali, di sicurezza, di regressione, di integrazione e di carico.
7. Risoluzione dei problemi relativi alle implementazioni.
8. La corretta convalida dei passaggi precedenti dovrebbe attivare un flusso di lavoro di approvazione manuale per autorizzare l'implementazione nell'ambiente di produzione.

Livello di impegno per il piano di implementazione: elevato

Risorse

Best practice correlate:

- OPS05-BP02 Testare e convalidare le modifiche

Documenti correlati:

- AWS Builders' Library | Automatizzazione di implementazioni sicure e pratiche | Distribuzioni di test
- AWS Whitepaper | Praticare l'integrazione continua e la distribuzione continua su AWS
- The Story of Apollo - Amazon's Deployment Engine
- Come eseguire test ed eseguire il debug AWS CodeDeploy localmente prima di spedire il codice
- Integrating Network Connectivity Testing with Infrastructure Deployment

Video correlati:

- re:Invent 2020 | Testing software and systems at Amazon

Esempi correlati:

- Tutorial | Implementazione e ECS servizio Amazon con un test di convalida

OPS06-BP03 Utilizzare strategie di implementazione sicure

I roll-out sicuri della produzione controllano il flusso di modifiche vantaggiose con l'obiettivo di ridurre al minimo l'impatto percepito di tali modifiche sui clienti. I controlli di sicurezza forniscono meccanismi di ispezione per convalidare i risultati desiderati e limitare l'ambito di impatto derivante da eventuali difetti introdotti dalle modifiche o da errori di implementazione. I roll-out sicuri possono includere strategie come feature-flags, one-box, roll-out (release canary), immutabili, suddivisioni del traffico e implementazioni blu/verdi.

Risultato desiderato: l'organizzazione utilizza un sistema di distribuzione e integrazione continua (CI/CD) che fornisce funzionalità per automatizzare roll-out sicuri. I team sono tenuti a utilizzare strategie di roll-out sicure appropriate.

Anti-pattern comuni:

- Implementi una modifica non riuscita a tutta la produzione contemporaneamente. Di conseguenza, tutti i clienti vengono colpiti contemporaneamente.
- Un difetto introdotto in un'implementazione simultanea su tutti i sistemi richiede una release di emergenza. La correzione per tutti i clienti richiede diversi giorni.
- La gestione della release di produzione richiede la pianificazione e la partecipazione di diversi team. Ciò limita la tua capacità di aggiornare frequentemente le funzionalità per i tuoi clienti.
- Esegui un'implementazione variabile modificando i sistemi esistenti. Dopo aver scoperto che la modifica non è andata a buon fine, devi modificare nuovamente i sistemi per ripristinare la versione precedente estendendo il tempo di ripristino.

Vantaggi dell'adozione di questa best practice: le implementazioni automatizzate bilanciano la velocità dei roll-out con la fornitura costante di modifiche vantaggiose per i clienti. La limitazione dell'impatto previene costosi errori di implementazione e massimizza la capacità dei team di rispondere in modo efficiente ai guasti.

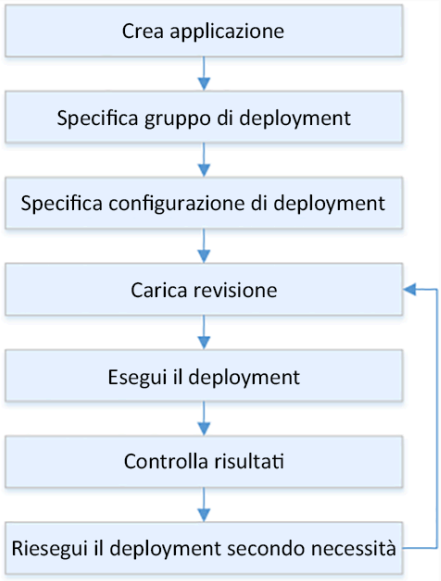
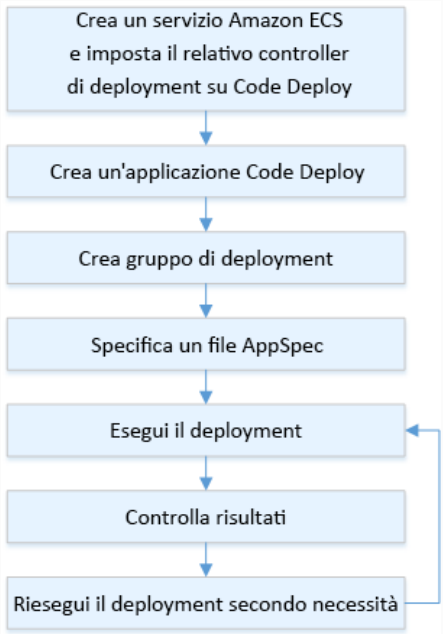
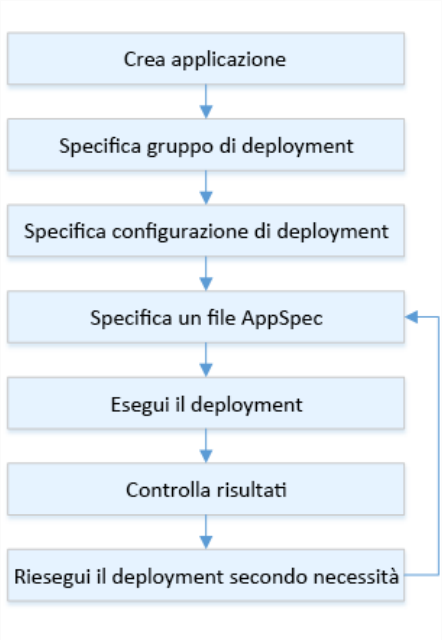
Livello di rischio associato se questa best practice non fosse adottata: medio

Guida all'implementazione

Gli errori della distribuzione continua possono portare a una ridotta disponibilità del servizio e a esperienze dei clienti negative. Per massimizzare la percentuale di implementazioni riuscite, implementate controlli di sicurezza nel processo di end-to-end rilascio per ridurre al minimo gli errori di implementazione, con l'obiettivo di azzerare gli errori di implementazione.

Esempio del cliente

AnyCompany L'obiettivo del settore retail è quello di ridurre al minimo o nullo le implementazioni con tempi di inattività, il che significa che non vi è alcun impatto percepibile sugli utenti durante le implementazioni. A tal fine, l'azienda ha stabilito modelli di implementazione (vedi il seguente diagramma del flusso di lavoro) come roll-out e implementazioni blu/verdi. Tutti i team adottano uno o più di questi modelli nella loro pipeline CI/CD.

CodeDeploy flusso di lavoro per Amazon EC2	CodeDeploy flusso di lavoro per Amazon ECS	CodeDeploy flusso di lavoro per Lambda
		

Passaggi dell'implementazione

1. Utilizza un flusso di lavoro di approvazione per avviare la sequenza delle fasi di roll-out della produzione al momento della promozione alla produzione.
2. Utilizza un sistema di distribuzione automatizzato come AWS CodeDeploy. AWS CodeDeploy le opzioni di distribuzione includono distribuzioni sul posto per EC2 /On-Premises e distribuzioni blu/green per /On-Premises e Amazon (vedi il EC2 diagramma del flusso di lavoro precedente AWS Lambda). ECS
   a. Ove applicabile, effettua l'integrazione con altri servizi o con prodotti e servizi dei partner. AWS CodeDeployAWSAWS CodeDeploy

3. Utilizza distribuzioni blu/verdi per database come Amazon Aurora e Amazon. RDS
4. Monitora le distribuzioni utilizzando le notifiche di eventi di Amazon e Amazon Simple Notification Service SNS (Amazon). CloudWatch AWS CloudTrail
5. Esegui test automatici post-implementazione, inclusi test funzionali, di sicurezza, di regressione, di integrazione e di carico.
6. Risoluzione dei problemi relativi alle implementazioni.

Livello di impegno per il piano di implementazione: medio

Risorse

Best practice correlate:

- OPS05-BP02 Testare e convalidare le modifiche
- OPS05-BP09 Apporta modifiche frequenti, piccole e reversibili
- OPS05- BP1 0 Integrazione e implementazione completamente automatizzate

Documenti correlati:

- AWS Builders Library | Automatizzazione di implementazioni pratiche e sicure | Implementazioni di produzione
- AWS Builders Library | La mia pipeline CI/CD è il mio release captain | Rilasci di produzione sicuri e automatici
- AWS Whitepaper | Praticare l'integrazione continua e la distribuzione continua su | Metodi di implementazione AWS
- AWS CodeDeploy Guida per l'utente
- Utilizzo delle configurazioni di distribuzione in AWS CodeDeploy
- Configura una distribuzione di API Gateway Canary Release
- Tipi ECS di distribuzione Amazon
- Implementazioni Blue/Green completamente gestite in Amazon Aurora e Amazon RDS
- Implementazioni blu/verdi con AWS Elastic Beanstalk

Video correlati:

- re:Invent 2020 | Hands-off: Automating continuous delivery pipelines at Amazon

- re:Invent 2019 | Amazon's Approach to high-availability deployment

Esempi correlati:

- Prova un esempio di implementazione Blue/Green in AWS CodeDeploy
- Workshop | Creazione di pipeline CI/CD per implementazioni Lambda Canary utilizzando AWS CDK
- Workshop | Implementazione di Blue/Green e EKS Canary per e ECS
- Workshop | Building a Cross-account CI/CD Pipeline

OPS06-BP04 Automatizza i test e il rollback

Per aumentare la velocità, l'affidabilità e la sicurezza del processo di implementazione, rendi disponibile una strategia per le funzionalità di test e rollback automatizzate negli ambienti di pre-produzione e produzione. Automatizza i test durante l'implementazione nella produzione per simulare le interazioni umane e di sistema che verificano le modifiche implementate. Automatizza il rollback per tornare rapidamente allo stato precedente corretto noto. Il rollback deve essere avviato automaticamente in condizioni predefinite, ad esempio quando il risultato desiderato della modifica non viene raggiunto o quando il test automatico fallisce. L'automazione di queste due attività migliora la percentuale di successo delle implementazioni, riduce al minimo i tempi di ripristino e riduce il potenziale impatto sulle attività aziendali.

Risultato desiderato: i test automatici e le strategie di rollback sono integrati nella pipeline di integrazione continua e distribuzione continua (CI/CD). Il monitoraggio è in grado di eseguire la convalida in base ai criteri di successo e avviare il rollback automatico in caso di errore. Ciò riduce al minimo l'impatto sugli utenti finali e sui clienti. Ad esempio, quando tutti i risultati dei test sono stati soddisfatti, promuovi il codice nell'ambiente di produzione in cui vengono avviati i test di regressione automatizzati, sfruttando gli stessi casi di test. Se i risultati dei test di regressione non corrispondono alle aspettative, viene avviato il rollback automatico nel flusso di lavoro della pipeline.

Anti-pattern comuni:

- I tuoi sistemi non sono progettati in modo da consentire loro di essere aggiornati con release più piccole. Di conseguenza, è difficile annullare tali modifiche di massa (bulk changes) durante un'implementazione conclusasi con esito negativo.
- Il processo di implementazione consiste in una serie di passaggi manuali. Dopo aver distribuito le modifiche al carico di lavoro, inizi i test post-implementazione. Dopo il test, ti rendi conto che il

tuo carico di lavoro è inutilizzabile e i clienti sono disconnessi. Inizi quindi a eseguire il rollback alla versione precedente. Tutti questi passaggi manuali ritardano il ripristino complessivo del sistema e provocano un impatto prolungato sui clienti.

- Hai impiegato del tempo a sviluppare casi di test automatizzati per funzionalità che non vengono utilizzate frequentemente nella tua applicazione, riducendo al minimo il ritorno sull'investimento nella tua capacità di eseguire test automatizzati.
- La versione è composta da applicazioni, infrastrutture, patch e aggiornamenti di configurazione indipendenti l'uno dall'altro. Tuttavia, è disponibile un'unica pipeline CI/CD che fornisce tutte le modifiche contemporaneamente. Un guasto in un componente obbliga a ripristinare tutte le modifiche, rendendo il rollback complesso e inefficiente.
- Il tuo team completa il lavoro di codifica nello sprint uno e inizia il lavoro dello sprint due, ma il tuo piano non includeva i test fino allo sprint tre. Come conseguenza, i test automatici hanno rivelato difetti dello sprint uno che dovevano essere risolti prima di poter avviare il test dei deliverable dello sprint due e l'intera release viene ritardata, rendendo inutili i test automatizzati.
- I casi di test di regressione automatizzati per la release di produzione sono completi, ma non stai monitorando lo stato del carico di lavoro. Poiché non è possibile verificare se il servizio è stato riavviato o meno, non sei sicuro se il rollback sia necessario o se sia già avvenuto.

Vantaggi dell'adozione di questa best practice: i test automatizzati aumentano la trasparenza del processo di verifica e la capacità di coprire più funzionalità in un periodo di tempo più breve. Testando e convalidando le modifiche nella produzione, è possibile identificare immediatamente i problemi. Il miglioramento della coerenza con strumenti di test automatizzati consente una migliore rilevazione dei difetti. Effettuando automaticamente il rollback alla versione precedente, l'impatto sui clienti viene ridotto al minimo. In ultima analisi, il rollback automatizzato ispira maggiore fiducia nelle capacità di implementazione riducendo l'impatto sulle attività aziendali. Nel complesso, queste funzionalità si riducono garantendo al contempo la qualità. time-to-delivery

Livello di rischio associato se questa best practice non fosse adottata: medio

Guida all'implementazione

Automatizza i test degli ambienti implementati per verificare che i risultati siano quelli desiderati. Automatizza il rollback a uno stato corretto noto quando non vengono raggiunti i risultati previsti, per ridurre al minimo il tempo di ripristino e gli errori causati dai processi manuali. Integra gli strumenti di test con il flusso di lavoro della pipeline per testare in modo coerente e ridurre al minimo gli input manuali. Dai priorità all'automazione dei casi di test, come quelli che mitigano i rischi maggiori e

devono essere testati frequentemente a ogni modifica. Inoltre, automatizza il rollback in base a condizioni specifiche predefinite nel tuo piano di test.

Passaggi dell'implementazione

1. Stabilisci un ciclo di vita di test per il tuo ciclo di vita di sviluppo che definisca ogni fase del processo di test, dalla pianificazione dei requisiti allo sviluppo dei test case, alla configurazione degli strumenti, ai test automatizzati e alla chiusura dei test case.
   a. Crea un approccio di test specifico per il carico di lavoro partendo dalla tua strategia di test complessiva.
   b. Prendi in considerazione una strategia di test continuo, laddove appropriato, durante tutto il ciclo di vita dello sviluppo.
2. Seleziona strumenti automatizzati per il test e il rollback in base ai requisiti aziendali e agli investimenti nella pipeline.
3. Decidi quali casi di test desideri automatizzare e quali devono essere eseguiti manualmente. Questi possono essere definiti in base alla priorità del valore aziendale della funzionalità testata. Allinea tutti i membri del team su questo piano e verifica la responsabilità per l'esecuzione di test manuali.
   a. Applica le funzionalità di test automatico a casi di test specifici che è opportuno automatizzare, come i casi ripetibili o eseguiti di frequente, quelli che richiedono attività ripetitive o quelli non più necessari per più configurazioni.
   b. Definisci gli script di automazione dei test e i criteri di successo nello strumento di automazione in modo da poter avviare l'automazione continua del flusso di lavoro quando casi specifici falliscono.
   c. Definisci criteri di errore specifici per il rollback automatico.
4. Dai priorità all'automazione dei test per ottenere risultati coerenti con lo sviluppo accurato e completo di casi di test in cui la complessità e l'interazione umana hanno un rischio maggiore di fallimento.
5. Integra i tuoi strumenti di test e rollback automatizzati nella tua pipeline CI/CD.
   a. Sviluppa criteri di successo chiari per le tue modifiche.
   b. Monitora e osserva per rilevare questi criteri e annullare automaticamente le modifiche quando vengono soddisfatti criteri di rollback specifici.
6. Esegui diversi tipi di test di produzione automatizzati, come:
   a. Test A/B, per mostrare i risultati rispetto alla versione corrente tra due gruppi di utenti di test.

   b. Test canary, che consente di distribuire la modifica a un sottoinsieme di utenti prima di rilasciarla a tutti.
   c. Test con flag delle funzionalità, che consente di attivare e disattivare una singola funzionalità della nuova versione alla volta dall'esterno dell'applicazione, in modo che ogni nuova funzionalità possa essere convalidata una alla volta.
   d. Test di regressione, per verificare nuove funzionalità con componenti correlati esistenti.
7. Monitora gli aspetti operativi dell'applicazione, delle transazioni e delle interazioni con altre applicazioni e componenti. Sviluppa report per mostrare il successo delle modifiche in base al carico di lavoro in modo da poter identificare quali parti dell'automazione e del flusso di lavoro possono essere ulteriormente ottimizzate.
   a. Sviluppa report sui risultati dei test che ti aiutino a prendere decisioni rapide sull'opportunità o meno di richiamare o meno le procedure di rollback.
   b. Implementa una strategia che consenta il rollback automatico basato su condizioni di errore predefinite derivanti da uno o più metodi di test.
8. Sviluppa i tuoi casi di test automatizzati per consentire la riutilizzabilità in caso di modifiche ripetibili future.

Livello di impegno per il piano di implementazione: medio

Risorse

Best practice correlate:

- OPS06-BP01 Piano per modifiche non riuscite
- OPS06-BP02 Implementazioni di test

Documenti correlati:

- AWS Builders Library | Garantire la sicurezza in caso di rollback durante le implementazioni
- Ridistribuisci e ripristina una distribuzione con AWS CodeDeploy
- 8 best practice per automatizzare le implementazioni con AWS CloudFormation

Esempi correlati:

- Test dell'interfaccia utente senza server utilizzando Selenium e Developer Tools AWS LambdaAWS FargateAWS

Video correlati:

- re:Invent 2020 | Hands-off: Automating continuous delivery pipelines at Amazon
- re:Invent 2019 | Amazon's Approach to high-availability deployment

OPS7. Come fai a sapere se hai tutto pronto per supportare un carico di lavoro?

Valuta la prontezza operativa del carico di lavoro, dei processi e delle procedure, nonché del personale per comprendere i rischi operativi correlati al carico di lavoro.

Best practice

- OPS07-BP01 Garantire la capacità del personale
- OPS07-BP02 Garantire una revisione coerente della prontezza operativa
- OPS07-BP03 Usa i runbook per eseguire le procedure
- OPS07-BP04 Usa i playbook per analizzare i problemi
- OPS07-BP05 Prendere decisioni informate per implementare sistemi e modifiche
- OPS07-BP06 Creare piani di supporto per carichi di lavoro di produzione

OPS07-BP01 Garantire la capacità del personale

Predisponi un meccanismo per stabilire se è disponibile il numero appropriato di risorse qualificate per supportare il carico di lavoro. Le risorse devono essere state formate sulla piattaforma e sui servizi che costituiscono il tuo carico di lavoro. Fornisci loro le informazioni necessarie per eseguire il carico di lavoro. Devi avere a disposizione personale qualificato sufficiente per supportare il normale funzionamento del carico di lavoro e gestire gli eventuali incidenti. Predisponi personale sufficiente per la rotazione durante la reperibilità e le ferie per evitare motivi di frustrazione.

Risultato desiderato:

- Presenza di personale qualificato sufficiente per supportare il carico di lavoro nei momenti in cui è disponibile.
- Capacità di fornire al personale formazione sul software e sui servizi che costituiscono il carico di lavoro.

Anti-pattern comuni:

- Implementazione di un carico di lavoro senza membri del team qualificati per l'esecuzione della piattaforma e dei servizi in uso.
- Mancanza di personale sufficiente per supportare la reperibilità a rotazione o le richieste di permesso del personale.

Vantaggi dell'adozione di questa best practice:

- Presenza di membri del team qualificati che offrono un supporto efficace al carico di lavoro.
- Con un numero sufficiente di membri del team, puoi supportare il carico di lavoro e la reperibilità a rotazione, riducendo il rischio di frustrazione.

Livello di rischio associato se questa best practice non fosse adottata: elevato

Guida all'implementazione

Verifica che sia disponibile personale qualificato sufficiente per supportare il carico di lavoro. Assicurati che il numero di membri del team di cui disponi sia sufficiente a coprire le normali attività operative, inclusa la reperibilità a rotazione.

Esempio del cliente

AnyCompany Retail si assicura che i team che supportano il carico di lavoro dispongano di personale e formazione adeguati. L'azienda ha al suo interno un numero sufficiente di tecnici per supportare la reperibilità a rotazione. Il personale riceve formazione sul software e sulla piattaforma su cui è basato il carico di lavoro e viene incoraggiato a conseguire certificazioni. Vi è personale sufficiente per permettere alle persone di richiedere permessi di assenza, continuando a supportare il carico di lavoro durante la reperibilità a rotazione.

Passaggi dell'implementazione

1. Assegna un numero adeguato di risorse del personale per eseguire e supportare il carico di lavoro, tenendo conto della reperibilità.
2. Forma il personale sul software e sulle piattaforme che costituiscono il carico di lavoro.
   a. AWS Training and Certification dispone di una libreria di corsi su. AWS Sono disponibili corsi gratuiti e a pagamento, online e di persona.
   b. AWS ospita eventi e webinar in cui impari dagli AWS esperti.

3. Valuta regolarmente le dimensioni e le competenze del team in base al mutare delle condizioni operative e del carico di lavoro. Adegua le dimensioni e le competenze del team ai requisiti operativi.

Livello di impegno per il piano di implementazione: elevato. L'assunzione e la formazione di un team per supportare il carico di lavoro possono richiedere un impegno significativo, ma assicurano solidi vantaggi a lungo termine.

Risorse

Best practice correlate:

- OPS11-BP04 Eseguire la gestione della conoscenza: i membri del team devono disporre delle informazioni necessarie per eseguire e supportare il carico di lavoro. La gestione delle informazioni è il fattore chiave a questo scopo.

Documenti correlati:

- AWS Eventi e webinar
- AWS Formazione e certificazione

OPS07-BP02 Garantire una revisione coerente della prontezza operativa

Utilizza Operational Readiness Reviews (ORRs) per verificare la capacità di gestire il tuo carico di lavoro. ORRè un meccanismo sviluppato da Amazon per verificare che i team possano gestire in sicurezza i propri carichi di lavoro. An ORR è un processo di revisione e ispezione che utilizza un elenco di requisiti. An ORR è un'esperienza self-service che i team utilizzano per certificare i propri carichi di lavoro. ORRsincludono le migliori pratiche tratte dalle lezioni apprese durante i nostri anni di sviluppo di software.

Una ORR lista di controllo è composta da raccomandazioni sull'architettura, sui processi operativi, sulla gestione degli eventi e sulla qualità dei rilasci. Il nostro processo di correzione dell'errore (CoE, Correction of Error) è uno dei principali fattori trainanti di questi elementi. La vostra analisi post-incidente dovrebbe guidare la vostra evoluzione. ORR An non ORR consiste solo nel seguire le migliori pratiche, ma anche nel prevenire il ripetersi di eventi già visti in precedenza. Infine, i requisiti di sicurezza, governance e conformità possono essere inclusi anche in un. ORR

Esegui ORRs prima che un carico di lavoro raggiunga la disponibilità generale e quindi durante tutto il ciclo di vita dello sviluppo del software. L'esecuzione ORR prima del lancio aumenta la capacità di gestire il carico di lavoro in sicurezza. Riesegui periodicamente ORR il carico di lavoro per catturare eventuali deviazioni dalle best practice. Puoi avere ORR liste di controllo per il lancio di nuovi servizi e per le revisioni periodiche. ORRs In tal modo puoi tenerti aggiornato sulle nuove best practice che emergono e incorporare le lezioni apprese dall'analisi post-incidente. Man mano che l'uso del cloud matura, puoi incorporare ORR requisiti predefiniti nella tua architettura.

Risultato desiderato: hai una ORR lista di controllo con le migliori pratiche per la tua organizzazione. ORRsvengono eseguiti prima del lancio dei carichi di lavoro. ORRsvengono eseguiti periodicamente nel corso del ciclo di vita del carico di lavoro.

Anti-pattern comuni:

- Avvii un carico di lavoro senza sapere se puoi utilizzarlo.
- I requisiti di governance e sicurezza non sono inclusi nella certificazione di un carico di lavoro per l'avvio.
- I carichi di lavoro non vengono rivalutati periodicamente.
- I carichi di lavoro vengono avviati senza le procedure richieste.
- Si osserva la ripetizione di errori con la stessa causa principale in più carichi di lavoro.

Vantaggi dell'adozione di questa best practice:

- I tuoi carichi di lavoro includono le best practice di architettura, processo e gestione.
- Le lezioni apprese vengono incorporate nel processo. ORR
- Le procedure richieste sono in atto all'avvio dei carichi di lavoro.
- ORRsvengono eseguiti per l'intero ciclo di vita del software dei carichi di lavoro.

Livello di rischio se questa best practice non fosse adottata: elevato

Guida all'implementazione

An ORR è composto da due cose: un processo e una lista di controllo. Il ORR processo deve essere adottato dall'organizzazione e supportato da uno sponsor esecutivo. Come minimo, ORRs deve essere eseguito prima che un carico di lavoro raggiunga la disponibilità generale. Esegui l'ORRintero ciclo di vita di sviluppo del software per mantenerlo aggiornato con le migliori pratiche o i nuovi

requisiti. La ORR lista di controllo dovrebbe includere elementi di configurazione, requisiti di sicurezza e governance e le migliori pratiche dell'organizzazione. Nel tempo, è possibile utilizzare servizi come, e AWS Control Tower Guardrails AWS ConfigAWS Security Hub, per sviluppare le migliori pratiche a partire da ORR guardrails per il rilevamento automatico delle migliori pratiche.

Esempio del cliente

Dopo diversi incidenti di produzione, AnyCompany Retail ha deciso di implementare un processo. ORR Ha creato un elenco di controllo composto da best practice, requisiti di governance e conformità e lezioni apprese dalle interruzioni. I nuovi carichi di lavoro vengono eseguiti ORRs prima del lancio. Ogni carico di lavoro viene eseguito annualmente ORR con un sottoinsieme di best practice per incorporare nuove best practice e requisiti che vengono aggiunti alla lista di controllo. ORR Nel corso del tempo, AnyCompany Retail individuava alcune best practice, AWS Configche velocizzavano il processo. ORR

Passaggi dell'implementazione

Per saperne di piùORRs, leggi il white paper Operational Readiness Reviews (ORR). Fornisce informazioni dettagliate sulla storia del ORR processo, su come costruire la propria ORR pratica e su come sviluppare la propria lista di controllo. ORR I passaggi seguenti costituiscono una versione abbreviata di quel documento. Per una comprensione approfondita di cosa ORRs sono e come crearne uno proprio, consigliamo di leggere il white paper.

1. Riunisci le parti interessate importanti, inclusi i rappresentanti della sicurezza, delle operazioni e dello sviluppo.
2. Chiedi a ogni parte interessata di indicare almeno un requisito. Per la prima iterazione, prova a limitare il numero di elementi a trenta al massimo.
   - Appendice B: ORR Le domande di esempio tratte dal white paper Operational Readiness Reviews (ORR) contengono esempi di domande che è possibile utilizzare per iniziare.
3. Raccogli i tuoi requisiti in un foglio di calcolo.
   - Puoi utilizzare obiettivi personalizzati per sviluppare i AWS Well-Architected Tooltuoi obiettivi ORR e condividerli tra i tuoi account e la tua organizzazione. AWS
4. Identifica un carico di lavoro a cui ORR dedicarti. L'ideale è un carico di lavoro pre-lancio o un carico di lavoro interno.
5. Consulta la ORR lista di controllo e prendi nota di tutte le scoperte fatte. I rilevamenti potrebbero non essere validi se è in atto una mitigazione. Aggiungi qualsiasi rilevamento privo di mitigazione al tuo backlog di elementi e implementalo prima del lancio.

6. Continua ad aggiungere le migliori pratiche e requisiti alla tua ORR lista di controllo nel tempo.

AWS Support i clienti con Enterprise Support possono richiedere l'Operational Readiness Review Workshop al proprio Technical Account Manager. Il workshop è una sessione interattiva di lavoro a ritroso per sviluppare la propria ORR lista di controllo.

Livello di impegno per il piano di implementazione: elevato. L'adozione di una ORR prassi all'interno dell'organizzazione richiede la sponsorizzazione dei dirigenti e il consenso delle parti interessate. Crea e aggiorna l'elenco di controllo con input provenienti da tutta l'organizzazione.

Risorse

Best practice correlate:

- OPS01-BP03 Valuta i requisiti di governance— I requisiti di governance sono una scelta naturale per una lista di controllo. ORR
- OPS01-BP04 Valuta i requisiti di conformità— I requisiti di conformità sono talvolta inclusi in una lista di ORR controllo. Altre volte costituiscono un processo separato.
- OPS03-BP07 Team di risorse appropriati— La capacità del team è un buon candidato per un ORR requisito.
- OPS06-BP01 Piano per modifiche non riuscite: prima di avviare il carico di lavoro, è necessario stabilire un piano di rollback o rollforward.
- OPS07-BP01 Garantire la capacità del personale: per supportare un carico di lavoro è necessario disporre del personale necessario.
- SEC01-BP03 Identifica e convalida gli obiettivi di controllo: gli obiettivi di controllo della sicurezza rappresentano requisiti eccellenti. ORR
- REL13-BP01 Definire gli obiettivi di ripristino per i tempi di inattività e la perdita di dati: i piani di disaster recovery sono un buon requisito. ORR
- COST02-BP01 Sviluppa politiche basate sui requisiti della tua organizzazione: è bene includere politiche di gestione dei costi nella tua lista di controllo. ORR

Documenti correlati:

- AWS Control Tower - Guardrail in AWS Control Tower
- AWS Well-Architected Tool - Lenti personalizzate

- Operational Readiness Review Template di Adrian Hornsby
- White paper sulle revisioni della prontezza operativa (ORR)

Video correlati:

- AWS Support S You | Creazione di un'efficace revisione della prontezza operativa () ORR

Esempi correlati:

- Esempio di revisione della prontezza operativa () ORR Lente

Servizi correlati:

- AWS Config
- AWS Control Tower
- AWS Security Hub
- AWS Well-Architected Tool

OPS07-BP03 Usa i runbook per eseguire le procedure

Un runbook è un processo documentato finalizzato al raggiungimento di un determinato risultato. I runbook sono composti da una serie di passaggi che è necessario eseguire per conseguire un obiettivo. L'uso dei runbook può essere fatto risalire agli albori dell'aviazione. Nelle operazioni cloud, è possibile utilizzare i runbook per ridurre i rischi e ottenere i risultati desiderati. In estrema sintesi, un runbook è un elenco di controllo da seguire per completare un'attività.

I runbook costituiscono una parte essenziale del funzionamento dei carichi di lavoro. Dall'onboarding di un nuovo membro in un team all'implementazione di una versione principale, i runbook sono processi codificati che garantiscono risultati coerenti indipendentemente da chi li utilizza. I runbook devono essere pubblicati a livello centralizzato e aggiornati in base all'evoluzione del processo. L'aggiornamento dei runbook rappresenta infatti un elemento chiave dell'intero processo di gestione delle modifiche. Devono inoltre includere le linee guida relative a gestione degli errori, strumenti, autorizzazioni, eccezioni ed escalation in caso di problemi.

Man mano che l'organizzazione cresce, è consigliabile automatizzare i runbook. Inizia con runbook concisi e di frequente utilizzo. Utilizza un linguaggio di scripting per automatizzare le procedure o

semplificarne l'esecuzione. Dopo aver automatizzato i primi runbook, potrai dedicare altro tempo all'automazione dei runbook più complessi. Gradualmente dovrai automatizzare la maggior parte dei runbook.

Risultato desiderato: il team dispone di una raccolta di step-by-step guide per l'esecuzione delle attività relative al carico di lavoro. I runbook contengono il risultato desiderato, gli strumenti e le autorizzazioni necessari e le istruzioni per la gestione degli errori. Vengono archiviati in una posizione centralizzata (sistema di controllo delle versioni) e aggiornati di frequente. Ad esempio, i runbook forniscono ai team funzionalità per monitorare, comunicare e rispondere agli AWS Health eventi per gli account critici durante gli allarmi delle applicazioni, i problemi operativi e gli eventi pianificati del ciclo di vita.

Anti-pattern comuni:

- Ricorso alla memoria per completare i singoli passaggi di un processo.
- Implementazione manuale delle modifiche senza utilizzare un elenco di controllo.
- Vari membri dei team eseguono lo stesso processo con procedure o risultati diversi.
- Mancato aggiornamento dei runbook in base alle modifiche o ai processi di automazione del sistema.

Vantaggi dell'adozione di questa best practice:

- Riduzione della percentuale degli errori per le attività manuali.
- Le operazioni vengono eseguite in modo coerente.
- I nuovi membri dei team possono essere operativi da subito.
- I runbook possono essere automatizzati per semplificare le operazioni più impegnative.

Livello di rischio associato se questa best practice non fosse adottata: medio

Guida all'implementazione

I runbook possono avere vari formati, a seconda del livello di "maturità" dell'organizzazione. Dovrebbero essere costituiti almeno da un documento di testo. step-by-step Il risultato desiderato deve essere indicato in modo chiaro e preciso. Devono inoltre documentare in modo chiaro le autorizzazioni e gli strumenti speciali necessari. Devono includere linee guida dettagliate relative alla gestione degli errori e ai livelli di escalation nel caso in cui si verifichino problemi o errori. I runbook

devono riportare il nome del proprietario ed essere pubblicati in una posizione centralizzata. Dopo averlo compilato, un runbook deve essere convalidato. A tale scopo, devi far predisporre il runbook da un membro diverso del tuo team. Con l'evoluzione della procedura, aggiorna i runbook in base al processo di gestione delle modifiche.

I runbook in formato testuale devono essere automatizzati a seconda dell'evoluzione dell'organizzazione. L'utilizzo di servizi come le automazioni di AWS Systems Manager ti consentono di trasformare un testo non formattato in automazioni che possono essere eseguite nell'ambito di un carico di lavoro. Queste automazioni possono essere eseguite in risposta agli eventi, riducendo l'onere operativo necessario per mantenere il carico di lavoro. AWS Systems Manager Automation offre anche un'esperienza di progettazione visiva a basso codice per creare più facilmente runbook di automazione.

Esempio del cliente

AnyCompany La vendita al dettaglio deve eseguire gli aggiornamenti dello schema del database durante le implementazioni del software. Il team responsabile delle operazioni cloud ha lavorato assieme al team addetto all'amministrazione del database per redigere un runbook per l'implementazione manuale di queste modifiche. Nel runbook sono incluse le procedure dettagliate sotto forma di elenco di controllo. È presente anche una sezione sulla gestione degli errori in caso di problemi. Il runbook è stato pubblicato assieme ad altri runbook sul wiki interno. Il team responsabile delle operazioni cloud pensa di pianificare l'automazione del runbook in futuro.

Passaggi dell'implementazione

Se non è presente un repository di documenti, è consigliabile creare una libreria di runbook utilizzando un repository per il controllo delle versioni. Puoi creare i runbook utilizzando Markdown. Di seguito è riportato un modello di runbook di esempio che è possibile utilizzare come riferimento per la creazione dei runbook.

```
# Runbook Title
## Runbook Info
| Runbook ID | Description | Tools Used | Special Permissions | Runbook Author | Last
 Updated | Escalation POC |
|-------|-------|-------|-------|-------|-------|-------|
| RUN001 | What is this runbook for? What is the desired outcome? | Tools | Permissions
 | Your Name | 2022-09-21 | Escalation Name |
## Steps
1. Step one
2. Step two
```

1. Se non disponi di un repository o di un wiki per la documentazione, crea un repository per il controllo delle versioni nel sistema di controllo delle versioni in uso.
2. Individua un processo che non ha un runbook. Un processo ideale viene eseguito a cadenza più o meno regolare, con un numero limitato di passaggi e con errori a basso impatto.
3. Nel repository di documenti, crea una nuova bozza di documento Markdown utilizzando il modello. Specifica il titolo del runbook e i campi obbligatori in Informazioni runbook.
4. Partendo dal primo passaggio, compila l'area Passaggi del runbook.
5. Associa il runbook a un membro del team. Chiedi a tale membro di utilizzare il runbook per convalidare i passaggi. In caso di informazioni mancanti o poca chiarezza, aggiorna il runbook.
6. Pubblica il runbook nell'archivio della documentazione interna. Comunica l'avvenuta pubblicazione al team e alle altre parti interessate.
7. In questo modo, nel corso del tempo creerai una libreria di runbook. Man mano che la libreria cresce, comincia a pensare di automatizzare i runbook.

Livello di impegno per il piano di implementazione: basso Lo standard minimo per un runbook è una step-by-step guida testuale. L'automazione dei runbook può aumentare l'impegno a livello di implementazione.

Risorse

Best practice correlate:

- OPS02-BP02 I processi e le procedure hanno identificato i proprietari
- OPS07-BP04 Usa i playbook per analizzare i problemi
- OPS10-BP01 Utilizza un processo per la gestione di eventi, incidenti e problemi
- OPS10-BP02 Predisponi di un processo per avviso
- OPS11-BP04 Esegui la gestione della conoscenza

Documenti correlati:

- Framework AWS Well-Architected, concetti: sviluppo di Runbook
- Achieving Operational Excellence using automated playbook and runbook
- AWS Systems Manager: utilizzo dei runbook
- Manuale di migrazione per migrazioni di AWS grandi dimensioni - Attività 4: Migliorare i runbook di migrazione

- Use AWS Systems Manager Automation runbooks to resolve operational tasks

Video correlati:

- AWS re:Invent 2019: DIY guida ai runbook, ai report sugli incidenti e alla risposta agli incidenti
- Come automatizzare le operazioni IT su AWS | Amazon Web Services
- Integrazione degli script in AWS Systems Manager

Esempi correlati:

- Well-Architected Labs: automazione delle operazioni con playbook e runbook
- AWS Post sul blog: Sviluppa una pratica di automazione del cloud per l'eccellenza operativa: le migliori pratiche di AWS Managed Services
- AWS Systems Manager: procedure dettagliate per l'automazione
- AWS Systems Manager: ripristina un volume root dall'ultimo runbook di snapshot
- Creazione di un runbook di risposta agli AWS incidenti utilizzando i notebook Jupyter e Lake CloudTrail
- Gitlab: runbook
- Rubix: una libreria Python per la creazione di runbook in notebook Jupyter
- Using Document Builder to create a custom runbook

Servizi correlati:

- AWS Systems Manager Automation

OPS07-BP04 Usa i playbook per analizzare i problemi

I playbook sono step-by-step guide utilizzate per indagare su un incidente. quando si verificano incidenti per analizzare, valutare l'impatto e identificare la causa principale del problema. I playbook sono utili in molti scenari diversi, dalle implementazioni non riuscite agli incidenti di sicurezza. In molti casi, i playbook identificano la causa principale che viene poi mitigata tramite un runbook. I playbook costituiscono un componente essenziale dei piani di risposta agli incidenti di ogni organizzazione.

Un buon playbook include diverse caratteristiche principali che guidano l'utente, passo dopo passo, nel processo di rilevamento. Ma quali passaggi deve eseguire l'utente per diagnosticare un incidente?

Illustra chiaramente nel playbook se sono necessari strumenti speciali o autorizzazioni elevate. È essenziale predisporre un piano di comunicazione per aggiornare le parti interessate sullo stato dell'analisi. Nelle situazioni in cui non è possibile identificare la causa principale, il playbook deve prevedere un piano di escalation. Se viene identificata la causa principale, il playbook deve includere il riferimento di un runbook che descrive come risolvere il problema. I playbook devono essere archiviati a livello centrale e aggiornati regolarmente. Se i playbook vengono utilizzati per avvisi specifici, fornisci al team i riferimenti dei playbook all'interno degli avvisi.

Man mano che l'organizzazione acquisisce maturità, puoi automatizzare i playbook. Inizia con i playbook che trattano incidenti a basso rischio. Utilizza gli script per automatizzare le procedure di rilevamento. Assicurati di avere i relativi runbook per mitigare le cause principali più comuni.

Risultato desiderato: la tua organizzazione dispone dei playbook per gli incidenti comuni. I playbook sono archiviati in una posizione centrale e disponibili per i membri del team. I playbook vengono aggiornati di frequente. Per qualsiasi causa principale nota, vengono creati i relativi runbook.

Anti-pattern comuni:

- Non esiste un modo standard per analizzare un incidente.
- I membri del team confidano nella "memoria muscolare" o nelle conoscenze istituzionali per risolvere i problemi di un'implementazione non riuscita.
- I nuovi membri del team apprendono come analizzare i problemi attraverso tentativi ed errori.
- Le best practice per l'analisi dei problemi non sono condivise tra i team.

Vantaggi dell'adozione di questa best practice:

- I playbook rendono più efficaci le tue attività di mitigazione degli incidenti.
- Uno stesso playbook può essere utilizzato da diversi membri del team in modo da identificare la causa principale in modo coerente.
- Le cause principali note possono già disporre di runbook appositamente sviluppati, accelerando i tempi di ripristino.
- I playbook contribuiscono ad accelerare la collaborazione tra i membri del team.
- I team possono applicare i processi su vasta scala tramite i playbook ripetibili.

Livello di rischio associato se questa best practice non fosse adottata: medio

Guida all'implementazione

Il modo in cui crei e utilizzi i playbook dipende dalla maturità della tua organizzazione. Se non hai familiarità con il cloud, crea i playbook in formato testo in un repository per i documenti centrale. Man mano che l'organizzazione acquisisce maturità, i playbook possono diventare semiautomatizzati tramite script scritti in linguaggi come Python. Questi script possono essere eseguiti all'interno di un notebook Jupyter per accelerare il rilevamento. Le organizzazioni avanzate dispongono di playbook completamente automatizzati per i problemi comuni che vengono risolti automaticamente con i runbook.

Inizia a creare i playbook elencando gli incidenti comuni che si verificano nel tuo carico di lavoro. Scegli i playbook per gli incidenti a basso rischio e in cui la causa principale è riconducibile a pochi problemi. Una volta creati i playbook per gli scenari più semplici, passa agli scenari a rischio più elevato o in cui la causa principale non è ancora nota.

I playbook in formato testo vengono automatizzati man mano che l'organizzazione acquisisce maturità. L'utilizzo di servizi come le automazioni di AWS Systems Manager ti consentono di trasformare un semplice testo in automazioni eseguibili sul carico di lavoro per accelerare le analisi. Queste automazioni possono essere attivate in risposta agli eventi, riducendo il tempo medio per rilevare e risolvere gli incidenti.

Grazie a AWS Systems Manager Incident Manager, i clienti possono rispondere agli incidenti. Questo servizio fornisce un'unica interfaccia per valutare gli incidenti, informare le parti interessate circa il rilevamento e la mitigazione e collaborare per tutta la durata dell'incidente. Utilizza AWS Systems Manager Automations per velocizzare il rilevamento e il ripristino.

Esempio del cliente

Un incidente di produzione ha avuto un impatto sulla vendita AnyCompany al dettaglio. L'ingegnere di turno utilizza un playbook per analizzare il problema e man mano che esegue i passaggi, mantiene aggiornate le parti interessate indicati nel playbook. L'ingegnere identifica la causa principale come una race condition di un servizio di backend. Utilizzando un runbook, l'ingegnere ha rilanciato il servizio, riportando Retail online. AnyCompany

Passaggi dell'implementazione

Se non è già presente, è consigliabile creare un repository per i documenti con il controllo delle versioni per la libreria di playbook. Puoi creare i tuoi playbook utilizzando Markdown, compatibile con la maggior parte dei sistemi di automazione dei playbook. Se parti da zero, utilizza il seguente modello di playbook come esempio.

```
# Playbook Title
## Playbook Info
| Playbook ID | Description | Tools Used | Special Permissions | Playbook Author | Last
 Updated | Escalation POC | Stakeholders | Communication Plan |
|-------|-------|-------|-------|-------|-------|-------|-------|-------|
| RUN001 | What is this playbook for? What incident is it used for? | Tools |
 Permissions | Your Name | 2022-09-21 | Escalation Name | Stakeholder Name | How will
 updates be communicated during the investigation? |
## Steps
1. Step one
2. Step two
```

1. Se non disponi di un repository o di un wiki per i documenti, crea un nuovo repository di controllo per il controllo delle versioni per i tuoi playbook nel tuo sistema di controllo delle versioni.
2. Identifica un problema comune che richieda un'analisi, vale a dire uno scenario in cui la causa principale è riconducibile a pochi problemi e la risoluzione è a basso rischio.
3. Utilizzando il modello Markdown, compila la sezione Titolo del playbook e i campi in Informazioni sul playbook.
4. Includi le procedure per la risoluzione dei problemi. Illustra nel modo più chiaro possibile le azioni da eseguire o le aree da analizzare.
5. Chiedi a un membro del team di esaminare e convalidare il tuo playbook. Se manca un'informazione o è necessario un chiarimento, aggiorna il playbook.
6. Pubblica il tuo playbook nel repository per i documenti e informa il tuo team e tutte le parti interessate.
7. Questa libreria di playbook diventerà sempre più ricca man mano che ne aggiungerai altri. Una volta che hai diversi playbook, inizia ad automatizzarli utilizzando strumenti come AWS Systems Manager Automations per mantenere sincronizzati automazione e playbook.

Livello di impegno per il piano di implementazione: basso I playbook sono documenti di testo archiviati in una posizione centrale. Le organizzazioni che hanno acquisito maturità applicano l'automazione dei playbook.

Risorse

Best practice correlate:

- OPS02-BP02 I processi e le procedure hanno identificato i proprietari

- OPS07-BP03 Usa i runbook per eseguire le procedure
- OPS10-BP01 Utilizza un processo per la gestione di eventi, incidenti e problemi
- OPS10-BP02 Predisponi di un processo per avviso
- OPS11-BP04 Esegui la gestione della conoscenza

Documenti correlati:

- Well-Architected AWS Framework, concetti: sviluppo di playbook
- Achieving Operational Excellence using automated playbook and runbook
- AWS Systems Manager: utilizzo dei runbook
- Use AWS Systems Manager Automation runbooks to resolve operational tasks

Video correlati:

- AWS re:Invent 2019: DIY guida ai runbook, ai report sugli incidenti e alla risposta agli incidenti (-R1) SEC318
- AWS Systems Manager Incident Manager - Workshop AWS virtuali
- Integrazione degli script in AWS Systems Manager

Esempi correlati:

- Framework AWS per playbook per i clienti
- AWS Systems Manager: procedure dettagliate per l'automazione
- Creazione di un runbook di risposta AWS agli incidenti utilizzando i notebook Jupyter e Lake CloudTrail
- Rubix: una libreria Python per la creazione di runbook in notebook Jupyter
- Using Document Builder to create a custom runbook
- Well-Architected Labs: automazione delle operazioni con playbook e runbook
- Well-Architected Labs: playbook di risposta agli incidenti con Jupyter

Servizi correlati:

- AWS Systems Manager Automation

- AWS Systems Manager Gestione incidenti

OPS07-BP05 Prendere decisioni informate per implementare sistemi e modifiche

Predisponi i processi per la gestione delle modifiche al carico di lavoro che hanno restituito esito positivo e negativo. Si definisce "pre-mortem" un esercizio in cui il team simula un errore per sviluppare strategie di mitigazione. Utilizza questo esercizio per prevedere errori e creare procedure ove opportuno. Valuta vantaggi e rischi dell'implementazione di modifiche nel carico di lavoro. Verifica che tutte le modifiche siano conformi ai requisiti di governance.

Risultato desiderato:

- Adozione di decisioni informate durante l'implementazione di modifiche nel carico di lavoro.
- Modifiche conformi ai requisiti di governance.

Anti-pattern comuni:

- Implementazione di una modifica nel carico di lavoro senza un processo per la gestione di un'implementazione errata.
- Applicazione di modifiche all'ambiente di produzione che non sono conformi ai requisiti di governance.
- Implementazione di una nuova versione del carico di lavoro senza stabilire valori di riferimento per l'utilizzo delle risorse.

Vantaggi dell'adozione di questa best practice:

- L'azienda è preparata all'effetto di modifiche infruttuose al carico di lavoro.
- Le modifiche apportate al carico di lavoro sono conformi ai criteri di governance.

Livello di rischio associato se questa best practice non fosse adottata: basso

Guida all'implementazione

Usa esercizi pre-mortem per sviluppare processi per la gestione di modifiche infruttuose. Documenta i processi di gestione delle modifiche infruttuose. Verifica che tutte le modifiche siano conformi ai requisiti di governance. Valuta vantaggi e rischi dell'implementazione di modifiche nel carico di lavoro.

Esempio del cliente

AnyCompany Retail effettua regolarmente verifiche preliminari per convalidare i propri processi in caso di modifiche non riuscite. L'azienda documenta i propri processi in un Wiki condiviso che aggiorna spesso. Tutte le modifiche sono conformi ai requisiti di governance.

Passaggi dell'implementazione

1. Prendi decisioni informate durante l'implementazione di modifiche nel carico di lavoro. Definisci ed esamina i criteri per un'implementazione corretta. Sviluppa scenari o criteri che avvierebbero il ripristino dello stato precedente a una modifica. Soppesa i vantaggi dell'implementazione di modifiche rispetto ai rischi di una modifica infruttuosa.
2. Verifica che tutte le modifiche siano conformi ai requisiti di governance.
3. Usa esercizi pre-mortem per pianificare la gestione delle modifiche infruttuose e documentare le strategie di mitigazione. Esegui un esercizio di simulazione di un'emergenza per modellare una modifica infruttuosa e convalidare le procedure di ripristino dello stato precedente.

Livello di impegno per il piano di implementazione: moderato L'implementazione di una procedura di pre-mortem richiede il coordinamento e l'impegno delle parti interessate in tutta l'organizzazione

Risorse

Best practice correlate:

- OPS01-BP03 Valuta i requisiti di governance: i requisiti di governance sono un fattore chiave per determinare se implementare una modifica.
- OPS06-BP01 Piano per modifiche non riuscite: predisponi piani per mitigare un'implementazione non riuscita e usa esercizi di pre-mortem per convalidarli.
- OPS06-BP02 Implementazioni di test: ogni modifica software deve essere testata nel modo adeguato prima dell'implementazione per ridurre gli errori nell'ambiente di produzione.
- OPS07-BP01 Garantire la capacità del personale: la presenza di personale qualificato sufficiente per supportare il carico di lavoro è essenziale per prendere una decisione informata riguardo all'implementazione di una modifica di sistema.

Documenti correlati:

- Amazon Web Services: rischio e conformità
- AWS Modello di responsabilità condivisa
- La governance nel mondo Cloud AWS: Il giusto equilibrio tra agilità e sicurezza

OPS07-BP06 Creare piani di supporto per carichi di lavoro di produzione

Abilita il supporto per qualsiasi software e servizio a cui si affida il tuo carico di lavoro di produzione. Seleziona un livello di supporto adeguato per soddisfare le esigenze di assistenza della produzione. I piani di supporto per queste dipendenze sono necessari nel caso si verifichi un'interruzione del servizio o un problema di software. Documenta i piani di supporto e come chiedere assistenza per tutti i servizi e i fornitori di software. Implementa meccanismi di verifica per controllare che i riferimenti del supporto siano aggiornati.

Risultato desiderato:

- Implementa piani di supporto per software e servizi a cui si affidano i carichi di lavoro di produzione.
- Scegli un piano di supporto adeguato in base alle esigenze di assistenza.
- Documenta i piani e i livelli di supporto e come richiedere assistenza.

Anti-pattern comuni:

- Non hai piani di supporto per un fornitore software strategico. Il tuo carico di lavoro ne risente e non puoi fare nulla per accelerare un intervento risolutivo o per ricevere aggiornamenti tempestivi dal fornitore.
- Uno sviluppatore, che era il punto di contatto primario di un fornitore di software, ha lasciato l'azienda. Non puoi contattare direttamente l'assistenza del fornitore. Devi investire il tuo tempo per cercare le informazioni e orientarti tra sistemi di contatto generici, aumentando così il livello di impegno richiesto per intervenire quando necessario.
- Si verifica un'interruzione della produzione con un fornitore di software. Non esiste una documentazione su come inserire una richiesta di assistenza.

Vantaggi dell'adozione di questa best practice:

- Con il livello di supporto adeguato, puoi ottenere una risposta nei tempi previsti per soddisfare le esigenze in termini di livelli di servizio.
- In caso di problemi in produzione, puoi effettuare l'escalation del problema se sei un cliente assistito.
- Fornitori di software e servizi possono essere di aiuto per la risoluzione dei problemi durante un incidente.

Livello di rischio associato se questa best practice non fosse adottata: basso

Guida all'implementazione

Abilita i piani di supporto per qualsiasi fornitore di software e servizi a cui si affida il tuo carico di lavoro di produzione. Configura piani di supporto adeguati per soddisfare le esigenze di assistenza. Per AWS i clienti, ciò significa attivare AWS Business Support o versioni successive su tutti gli account in cui sono presenti carichi di lavoro di produzione. Incontra con regolarità i fornitori del servizio di assistenza per ricevere aggiornamenti sulle offerte di supporto, sui processi e sui contatti. Documenta come richiedere assistenza ai fornitori di software e servizi, incluso come inoltrare il problema in caso si verificasse un'interruzione. Implementa meccanismi di aggiornamento dei contatti del supporto.

Esempio del cliente

In AnyCompany Retail, tutte le dipendenze di software e servizi commerciali prevedono piani di supporto. Ad esempio, hanno attivato AWS Enterprise Support su tutti gli account con carichi di lavoro di produzione. In caso di problemi, qualsiasi sviluppatore può inserire una richiesta di assistenza. Esiste una pagina wiki con informazioni su come richiedere assistenza, chi contattare e quali best practice seguire per accelerare il processo di risoluzione.

Passaggi dell'implementazione

1. Collabora con le parti interessate all'interno della tua organizzazione per identificare i fornitori di software e servizi su cui si basa il tuo carico di lavoro. Documenta queste dipendenze.
2. Stabilisci le esigenze in termini di assistenza del tuo carico di lavoro. Seleziona un piano di supporto in linea con tali esigenze.
3. Per software e servizi commerciali definisci un piano di supporto con i fornitori.
    a. L'abbonamento a AWS Business Support o superiore per tutti gli account di produzione offre tempi di risposta più rapidi AWS Support ed è fortemente consigliato. Se non disponi di un'assistenza premium, devi disporre di un piano d'azione per gestire i problemi, che richiedono l'assistenza di AWS Support. AWS Support offre un mix di strumenti e tecnologie, persone e programmi progettati per aiutarti in modo proattivo a ottimizzare le prestazioni, ridurre i costi e innovare più rapidamente. AWS Business Support offre vantaggi aggiuntivi, tra cui l'accesso alla AWS Trusted Advisor AWS Personal Health Dashboard e tempi di risposta più rapidi.
4. Documenta il tuo piano di supporto nello strumento di gestione delle conoscenze. Includi come richiedere assistenza, chi avvertire se viene inviata una richiesta di assistenza e come inoltrare il problema durante un incidente. Un wiki è un buon meccanismo che consente a tutti di apportare

gli aggiornamenti necessari alla documentazione, nel momento in cui vengono a conoscenza di modifiche a processi o contatti del supporto.

Livello di impegno per il piano di implementazione: basso La maggior parte di fornitori di servizi e software offre piani di supporto da attivare. Documentando e condividendo le best practice di supporto sul tuo sistema di gestione delle conoscenze, puoi verificare che il tuo team sappia cosa fare quando si verifica un problema in produzione.

Risorse

Best practice correlate:

- OPS02-BP02 I processi e le procedure hanno identificato i proprietari

Documenti correlati:

- AWS Support Piani

Servizi correlati:

- AWS Supporto aziendale
- AWS Supporto aziendale

Gestione

Questions

- OPS8. Come utilizzi l'osservabilità del carico di lavoro nella tua organizzazione?
- OPS9. Come fai a comprendere lo stato delle operazioni?
- OPS10. In che modo gestisci gli eventi del carico di lavoro e delle operazioni?

OPS8. Come utilizzi l'osservabilità del carico di lavoro nella tua organizzazione?

Garantire l'integrità del carico di lavoro sfruttando l'osservabilità. Utilizzare metriche, log e tracce pertinenti per ottenere una visione completa delle prestazioni del carico di lavoro e risolvere i problemi in modo efficiente.

Best practice

- OPS08-BP01 Analizza le metriche del carico di lavoro
- OPS08-BP02 Analizza i registri dei carichi di lavoro
- OPS08-BP03 Analizza le tracce del carico di lavoro
- OPS08-BP04 Crea avvisi utilizzabili
- OPS08-BP05 Crea dashboard

OPS08-BP01 Analizza le metriche del carico di lavoro

Dopo aver implementato la telemetria dell'applicazione, analizza regolarmente le metriche raccolte. Sebbene latenza, richieste, errori e capacità (o quote) forniscano informazioni dettagliate sulle prestazioni del sistema, è fondamentale dare priorità alla revisione delle metriche relative ai risultati aziendali. Ciò ti assicura di prendere decisioni basate sui dati in linea con i tuoi obiettivi aziendali.

Risultato desiderato: informazioni dettagliate sulle prestazioni del carico di lavoro che guidano decisioni basate sui dati, garantendo l'allineamento con gli obiettivi aziendali.

Anti-pattern comuni:

- Analisi isolata delle metriche senza considerare il loro impatto sui risultati aziendali.
- Eccessiva dipendenza dalle metriche tecniche trascurando quelle aziendali.
- Revisione poco frequente delle metriche, perdita di opportunità di prendere decisioni in tempo reale.

Vantaggi dell'adozione di questa best practice:

- Comprensione migliorata della correlazione tra prestazioni tecniche e risultati aziendali.
- Processo decisionale migliorato basato su dati in tempo reale.
- Identificazione e mitigazione proattive dei problemi prima che influiscano sui risultati aziendali.

Livello di rischio associato se questa best practice non fosse adottata: medio

Guida all'implementazione

Sfrutta strumenti come Amazon CloudWatch per eseguire analisi metriche. AWS servizi come il rilevamento delle CloudWatch anomalie e Amazon DevOps Guru possono essere utilizzati per

rilevare anomalie, soprattutto quando le soglie statiche sono sconosciute o quando i modelli di comportamento sono più adatti al rilevamento delle anomalie.

Passaggi dell'implementazione

1. Analizza e revisiona: revisiona e interpreta regolarmente le metriche relative al carico di lavoro.
   a. Dai priorità alle metriche relative ai risultati aziendali rispetto a quelle puramente tecniche.
   b. Comprendi l'importanza di picchi, cali o schemi nei dati.
2. Utilizza Amazon CloudWatch: utilizza Amazon CloudWatch per una visualizzazione centralizzata e un'analisi approfondita.
   a. Configura le CloudWatch dashboard per visualizzare le tue metriche e confrontarle nel tempo.
   b. Usa i percentili CloudWatch per avere una visione chiara della distribuzione delle metriche, che può aiutarti a definire e comprendere i valori anomali. SLAs
   c. Imposta il rilevamento delle CloudWatch anomalie per identificare modelli insoliti senza fare affidamento su soglie statiche.
   d. Implementa l'osservabilità CloudWatch tra più account per monitorare e risolvere i problemi delle applicazioni che si estendono su più account all'interno di una regione.
   e. Utilizza CloudWatch Metric Insights per interrogare e analizzare i dati metrici tra account e regioni, identificando tendenze e anomalie.
   f. CloudWatch Applica Metric Math per trasformare, aggregare o eseguire calcoli sulle tue metriche per ottenere informazioni più approfondite.
3. Utilizza Amazon DevOps Guru: incorpora Amazon DevOps Guru per il suo rilevamento delle anomalie potenziato dall'apprendimento automatico per identificare i primi segnali di problemi operativi per le tue applicazioni serverless e risolverli prima che abbiano un impatto sui tuoi clienti.
4. Ottimizza in base agli approfondimenti: prendi decisioni informate sulla base dell'analisi delle metriche per adeguare e migliorare i carichi di lavoro.

Livello di impegno per il piano di implementazione: medio

Risorse

Best practice correlate:

- OPS04-BP01 Identifica gli indicatori chiave di prestazione
- OPS04-BP02 Implementare la telemetria delle applicazioni

Documenti correlati:

- The Wheel Blog - Emphasizing the importance of continually reviewing metrics
- Percentile are important
- Usando AWS Cost Anomaly Detection
- CloudWatch osservabilità tra più account
- Interroga le tue metriche con Metrics Insights CloudWatch

Video correlati:

- Abilita l'osservabilità tra account in Amazon CloudWatch
- Introduzione ad Amazon DevOps Guru
- Analizza continuamente le metriche utilizzando AWS Cost Anomaly Detection

Esempi correlati:

- One Observability Workshop
- Acquisire informazioni operative AIOps con Amazon DevOps Guru

OPS08-BP02 Analizza i registri dei carichi di lavoro

L'analisi regolare dei log dei carichi di lavoro è essenziale per acquisire una comprensione più approfondita degli aspetti operativi dell'applicazione. Attraverso l'analisi, la consultazione e l'interpretazione efficiente dei dati di log, è possibile ottimizzare continuamente le prestazioni e la sicurezza delle applicazioni.

Risultato desiderato: informazioni dettagliate sul comportamento dell'applicazione e sulle operazioni derivanti da un'analisi completa dei log, che garantisce la rilevazione e la mitigazione proattiva dei problemi.

Anti-pattern comuni:

- Si trascura l'analisi dei log fino a quando non si verifica un problema critico.
- Il mancato utilizzo della suite completa degli strumenti disponibili per l'analisi dei log comporta la perdita di approfondimenti importanti.

- Si fa affidamento esclusivamente sulla revisione manuale dei log senza sfruttare le funzionalità di automazione e query.

Vantaggi dell'adozione di questa best practice:

- Identificazione proattiva dei colli di bottiglia operativi, delle minacce alla sicurezza e di altri problemi potenziali.
- Utilizzo efficiente dei dati di log per l'ottimizzazione continua dell'applicazione.
- Comprensione migliorata del comportamento dell'applicazione, facilitando il debug e la risoluzione dei problemi.

Livello di rischio associato se questa best practice non fosse adottata: medio

Guida all'implementazione

Amazon CloudWatch Logs è un potente strumento per l'analisi dei log. Funzionalità integrate come CloudWatch Logs Insights e Contributor Insights rendono il processo di derivazione di informazioni significative dai log intuitivo ed efficiente.

Passaggi dell'implementazione

1. Configurazione dei CloudWatch registri: configura applicazioni e servizi per inviare i log ai registri. CloudWatch
2. Usa il rilevamento delle anomalie nei log: utilizza il rilevamento delle anomalie di Amazon CloudWatch Logs per identificare e segnalare automaticamente modelli di log insoliti. Questo strumento consente di gestire in modo proattivo le anomalie nei log e di rilevare tempestivamente i potenziali problemi.
3. Configura CloudWatch Logs Insights: usa CloudWatch Logs Insights per cercare e analizzare in modo interattivo i tuoi dati di log.
   a. Crea query per estrarre modelli, visualizzare i dati di log e ricavare approfondimenti utili.
   b. Usa l'analisi dei pattern CloudWatch di Logs Insights per analizzare e visualizzare i pattern di log frequenti. Questa funzionalità consente di comprendere le tendenze operative più comuni e i potenziali valori anomali nei dati di log.
   c. Usa CloudWatch Logs compare (diff) per eseguire analisi differenziali tra diversi periodi di tempo o tra diversi gruppi di log. Questa funzionalità ti consente di individuare le modifiche e valutarne l'impatto sulle prestazioni o sul comportamento del sistema.

4. Monitora i log in tempo reale con Live Tail: usa Amazon CloudWatch Logs Live Tail per visualizzare i dati dei log in tempo reale. Puoi monitorare attivamente le attività operative dell'applicazione man mano che si verificano, ottenendo una visibilità immediata sulle prestazioni del sistema e sui potenziali problemi.
5. Sfrutta Contributor Insights: utilizza CloudWatchContributor Insights per identificare i migliori oratori in dimensioni ad alta cardinalità come gli indirizzi IP o gli user-agent.
6. Implementa i filtri metrici CloudWatch Logs: configura i filtri metrici CloudWatch Logs per convertire i dati di log in metriche utilizzabili. In questo modo puoi impostare allarmi o analizzare ulteriormente i modelli.
7. Implementa l'osservabilità CloudWatch tra account: monitora e risolvi i problemi delle applicazioni che si estendono su più account all'interno di una regione.
8. Rivedi regolarmente e perfeziona: rivedi periodicamente le tue strategie di analisi dei log per acquisire tutte le informazioni pertinenti e ottimizzare continuamente le prestazioni delle applicazioni.

Livello di impegno per il piano di implementazione: medio

Risorse

Best practice correlate:

- OPS04-BP01 Identifica gli indicatori chiave di prestazione
- OPS04-BP02 Implementare la telemetria delle applicazioni
- OPS08-BP01 Analizza le metriche del carico di lavoro

Documenti correlati:

- Analisi dei dati di registro con Logs Insights CloudWatch
- Utilizzo di Contributor Insights CloudWatch
- Creazione e gestione di filtri CloudWatch metrici di log

Video correlati:

- Analizza i dati di log con CloudWatch Logs Insights
- Usa CloudWatch Contributor Insights per analizzare dati ad alta cardinalità

Esempi correlati:

- CloudWatch Registra interrogazioni di esempio
- One Observability Workshop

OPS08-BP03 Analizza le tracce del carico di lavoro

L'analisi dei dati di tracciamento è fondamentale per ottenere una visione completa del percorso operativo di un'applicazione. Visualizzando e comprendendo le interazioni tra i vari componenti, consente di ottimizzare le prestazioni, identificare i colli di bottiglia e migliorare l'esperienza utente.

Risultato desiderato: ottieni una chiara visibilità sulle operazioni distribuite della tua applicazione, che si traduce in una risoluzione più rapida dei problemi e in un'esperienza utente migliorata.

Anti-pattern comuni:

- I dati di tracciamento vengono trascurati e ci si affida esclusivamente a log e metriche.
- I dati di tracciamento non sono correlati ai log associati.
- Vengono ignorate le metriche derivate dalle tracce, come la latenza e i tassi di errore.

Vantaggi dell'adozione di questa best practice:

- Migliora la risoluzione dei problemi e riduci il tempo medio di risoluzione (). MTTR
- Informazioni dettagliate sulle dipendenze e sul loro impatto.
- Identificazione e correzione rapide dei problemi legati alle prestazioni.
- Vengono sfruttate le metriche derivate dalle tracce per un processo decisionale informato.
- Esperienze utente migliorate attraverso interazioni con i componenti ottimizzate.

Livello di rischio associato se questa best practice non fosse adottata: medio

Guida all'implementazione

AWS X-Ray offre una suite completa per l'analisi dei dati di tracciamento, fornendo una visione olistica delle interazioni con i servizi, monitorando le attività degli utenti e rilevando i problemi di prestazioni. Funzionalità come X-Ray Insights ServiceLens, X-Ray Analytics e Amazon DevOps Guru migliorano la profondità delle informazioni fruibili derivate dai dati di tracciamento.

Passaggi dell'implementazione

I seguenti passaggi offrono un approccio strutturato per implementare efficacemente l'analisi dei dati di traccia utilizzando i servizi: AWS

1. Integrazione AWS X-Ray: assicurati che X-Ray sia integrato con le tue applicazioni per acquisire dati di traccia.
2. Analizza le metriche di X-Ray: approfondisci le metriche ottenute dalle tracce di X-Ray, come latenza, tassi di richieste, tassi di errore e distribuzioni dei tempi di risposta, utilizzando la mappa dei servizi per il monitoraggio dello stato delle applicazioni.
3. Utilizzo ServiceLens: sfrutta la ServiceLensmappa per una migliore osservabilità dei tuoi servizi e delle tue applicazioni. Fornisce la visualizzazione integrata di tracce, metriche, log, allarmi e altre informazioni correlate all'integrità.
4. Abilita X-Ray Insights:
   a. Attiva X-Ray Insights per rilevare in automatico le anomalie nelle tracce.
   b. Esamina gli approfondimenti per individuare i modelli e determinare le cause ultime, come l'aumento dei tassi di errore o delle latenze.
   c. Consulta la cronologia degli approfondimenti per un'analisi cronologica dei problemi rilevati.
5. Usa X-Ray Analytics: X-Ray Analytics ti consente di approfondire i dati di tracciamento, individuare modelli ed estrarre informazioni dettagliate.
6. Usa i gruppi di X-Ray: crea gruppi in X-Ray per filtrare le tracce in base a criteri come l'elevata latenza, per un'analisi più mirata.
7. Incorpora Amazon DevOps Guru: coinvolgi Amazon DevOps Guru per trarre vantaggio dai modelli di apprendimento automatico che individuano le anomalie operative nelle tracce.
8. Usa CloudWatch Synthetics: Usa Synthetics per creare CloudWatchcanarie per il monitoraggio continuo degli endpoint e dei flussi di lavoro. Questi canary possono integrarsi con X-Ray per fornire dati di tracciamento per un'analisi approfondita delle applicazioni testate.
9. Usa Real User Monitoring (RUM): con AWS X-Ray and CloudWatch RUM, puoi analizzare ed eseguire il debug del percorso della richiesta partendo dagli utenti finali della tua applicazione fino ai servizi gestiti a valle. AWS In questo modo, puoi identificare le tendenze e gli errori di latenza che hanno un impatto sugli utenti finali.
10. Effettua le correlazioni con i log: correla i dati di tracciamento con i log correlati all'interno della relativa vista di X-Ray per una prospettiva granulare sul comportamento delle applicazioni. Ciò consente di visualizzare gli eventi del log associati direttamente alle transazioni tracciate.

11. Implementa l'osservabilità CloudWatch tra account: monitora e risolvi i problemi delle applicazioni che si estendono su più account all'interno di una regione.

Livello di impegno per il piano di implementazione: medio

Risorse

Best practice correlate:

- OPS08-BP01 Analizza le metriche del carico di lavoro
- OPS08-BP02 Analizza i registri dei carichi di lavoro

Documenti correlati:

- Utilizzo ServiceLens per monitorare l'integrità delle applicazioni
- Esplorazione dei dati delle tracce con X-Ray Analytics
- Individuazione delle anomalie nelle tracce con X-Ray Insights
- Monitoraggio continuo con CloudWatch Synthetics

Video correlati:

- Analizza ed esegui il debug di applicazioni con Amazon CloudWatch Synthetics & AWS X-Ray
- Use AWS X-Ray Insights

Esempi correlati:

- One Observability Workshop
- Implementazione di X-Ray con AWS Lambda
- CloudWatchModelli Synthetics Canary

OPS08-BP04 Crea avvisi utilizzabili

Rilevare e rispondere tempestivamente alle deviazioni di comportamento dell'applicazione è fondamentale. Particolarmente importante è riconoscere quando i risultati basati sugli indicatori chiave di prestazione (KPIs) sono a rischio o quando si verificano anomalie impreviste. La base degli avvisi KPIs garantisce che i segnali ricevuti siano direttamente collegati all'impatto aziendale o

operativo. Questo approccio verso avvisi fruibili promuove risposte proattive e aiuta a mantenere le prestazioni e l'affidabilità del sistema.

Risultato desiderato: ricevi avvisi tempestivi, pertinenti e utilizzabili per identificare e mitigare rapidamente i potenziali problemi, soprattutto quando i risultati sono a rischio. KPI

Anti-pattern comuni:

- Si impostano troppi avvisi non critici, con conseguente affaticamento da avvisi ("alert fatigue").
- Non si dà priorità agli avvisi in base aKPIs, il che rende difficile comprendere l'impatto aziendale dei problemi.
- Non si affrontano le cause principali porta a ricevere avvisi ripetuti per lo stesso problema.

Vantaggi dell'adozione di questa best practice:

- Riduzione dell'affaticamento da avvisi ("alert fatigue") concentrandosi su avvisi pertinenti e fruibili.
- Maggiore operatività e affidabilità del sistema grazie al rilevamento e alla mitigazione proattiva dei problemi.
- Migliore collaborazione tra team e risoluzione più rapida dei problemi grazie all'integrazione con i più diffusi strumenti di avviso e comunicazione.

Livello di rischio associato se questa best practice non fosse adottata: elevato

Guida all'implementazione

Per creare un meccanismo di avviso efficace, è fondamentale utilizzare metriche, log e dati di tracciamento che segnalino quando i risultati basati su KPIs sono a rischio o vengono rilevate anomalie.

Passaggi dell'implementazione

1. Determina gli indicatori chiave di prestazione (KPIs): identifica quelli della tua applicazione. KPIs Gli avvisi devono essere collegati a questi KPIs per riflettere accuratamente l'impatto aziendale.
2. Implementa il rilevamento delle anomalie:
   - Usa il rilevamento delle CloudWatch anomalie di Amazon: configura il rilevamento delle CloudWatch anomalie di Amazon per rilevare automaticamente modelli insoliti, il che ti aiuta a generare avvisi solo per anomalie autentiche.
   - AWS X-Ray Usa Insights:

   a. Configura X-Ray Insights per la rilevazione delle anomalie nei dati di tracciamento.
   b. Configura le notifiche per X-Ray Insights così da ricevere avvisi sui problemi rilevati.
- Integrazione con Amazon DevOps Guru:
   a. Sfrutta Amazon DevOps Guru per le sue capacità di machine learning nel rilevare anomalie operative con i dati esistenti.
   b. Accedi alle impostazioni di notifica in DevOps Guru per configurare avvisi di anomalia.

3. Implementa avvisi fruibili: progetta avvisi che forniscano informazioni adeguate per intraprendere un'azione immediata.
   1. Monitora AWS Health gli eventi con EventBridge le regole di Amazon o esegui l'integrazione programmatica con le AWS Health API per automatizzare le azioni quando ricevi AWS Health eventi. Può trattarsi di azioni generali, come l'invio di tutti i messaggi pianificati sugli eventi del ciclo di vita a un'interfaccia di chat, oppure azioni specifiche, come l'avvio di un flusso di lavoro in uno strumento di gestione dei servizi IT.
4. Riduci l'affaticamento dagli avvisi: riduci al minimo gli avvisi non critici. Quando i team sono sovraccaricati da numerosi avvisi insignificanti, possono trascurare i problemi critici, riducendo l'efficacia complessiva del meccanismo di avviso.
5. Configurazione di allarmi compositi: utilizza gli allarmi CloudWatch compositi di Amazon per consolidare più allarmi.
6. Integrazione con strumenti di avviso: incorpora strumenti come Ops Genie e. PagerDuty
7. Coinvolgi AWS Chatbot: integra AWS Chatbotper inoltrare avvisi ad Amazon Chime, Microsoft Teams e Slack.
8. Avvisi basati sui log: utilizza i filtri metrici di log CloudWatch per creare allarmi basati su eventi di registro specifici.
9. Rivedi e itera: riesamina e ottimizza regolarmente le configurazioni degli avvisi.

Livello di impegno per il piano di implementazione: medio

Risorse

Best practice correlate:

- OPS04-BP01 Identifica gli indicatori chiave di prestazione
- OPS04-BP02 Implementare la telemetria delle applicazioni
- OPS04-BP03 Implementare la telemetria dell'esperienza utente

- OPS04-BP04 Implementare la telemetria delle dipendenze
- OPS04-BP05 Implementare la tracciabilità distribuita
- OPS08-BP01 Analizza le metriche del carico di lavoro
- OPS08-BP02 Analizza i registri dei carichi di lavoro
- OPS08-BP03 Analizza le tracce del carico di lavoro

Documenti correlati:

- Utilizzo degli CloudWatch allarmi Amazon
- Create a composite alarm
- Crea un CloudWatch allarme basato sul rilevamento delle anomalie
- DevOpsNotifiche Guru
- Notifiche X-Ray Insights
- Monitora, gestisci e risolvi i problemi delle tue AWS risorse con funzionalità interattive ChatOps
- Guida CloudWatch all'integrazione di Amazon | PagerDuty
- Integra Opsgenie con Amazon CloudWatch

Video correlati:

- Crea allarmi compositi in Amazon CloudWatch
- AWS Chatbot Panoramica
- AWS On Air ft. Comandi mutativi in AWS Chatbot

Esempi correlati:

- Allarmi, gestione degli incidenti e correzione nel cloud con Amazon CloudWatch
- Tutorial: creazione di una EventBridge regola Amazon che invia notifiche a AWS Chatbot
- One Observability Workshop

OPS08-BP05 Crea dashboard

Le dashboard rappresentano la visualizzazione incentrata sull'utente dei dati di telemetria dei carichi di lavoro. Sebbene forniscano un'interfaccia visiva fondamentale, non dovrebbero sostituire i meccanismi di allarme, ma integrarli. Se realizzate con cura, sono in grado di fornire approfondimenti

rapidi sullo stato e sulle prestazioni del sistema e possono informare le parti interessate in tempo reale riguardo ai risultati aziendali e all'impatto dei problemi.

Risultato desiderato:

Approfondimenti chiari e fruibili sullo stato del sistema e dell'azienda attraverso rappresentazioni visive.

Anti-pattern comuni:

- Dashboard eccessivamente complicate con troppe metriche.
- Affidarsi a dashboard senza avvisi per il rilevamento delle anomalie.
- Non aggiornare le dashboard man mano che i carichi di lavoro si evolvono.

Vantaggi di questa best practice:

- Visibilità immediata sulle metriche critiche del sistema e. KPIs
- Miglioramento della comunicazione e della comprensione con le parti interessate.
- Approfondimenti rapidi sull'impatto dei problemi operativi.

Livello di rischio associato se questa best practice non fosse adottata: medio

Guida all'implementazione

Dashboard incentrate sull'azienda

Le dashboard personalizzate per le aziende KPIs coinvolgono una gamma più ampia di parti interessate. Anche se queste persone potrebbero non essere interessate alle metriche di sistema, desiderano comprendere le implicazioni aziendali di questi numeri. Una dashboard incentrata sull'azienda garantisce che tutte le metriche tecniche e operative monitorate e analizzate siano allineate con gli obiettivi aziendali generali. Questo allineamento fornisce chiarezza, garantendo che tutti siano sulla stessa lunghezza d'onda per quanto riguarda ciò che è essenziale e ciò che non lo è. Inoltre, le dashboard che mettono in risalto il business KPIs tendono ad essere più fruibili. Le parti interessate possono comprendere rapidamente lo stato delle operazioni, le aree che richiedono attenzione e il potenziale impatto sui risultati aziendali.

Tenendo presente questo aspetto, quando crei le dashboard, assicurati che vi sia un equilibrio tra metriche tecniche e business. KPIs Entrambi sono fondamentali, ma si rivolgono a un pubblico diverso. Idealmente, dovresti disporre di dashboard che forniscano una visione olistica dello stato e

delle prestazioni del sistema, mettendo in evidenza al contempo i principali risultati aziendali e le loro implicazioni.

Le CloudWatch dashboard di Amazon sono home page personalizzabili nella CloudWatch console che puoi utilizzare per monitorare le tue risorse in un'unica visualizzazione, anche quelle distribuite su diversi Regioni AWS account.

Passaggi dell'implementazione

1. Crea una dashboard di base: crea una nuova dashboard in CloudWatch, assegnandole un nome descrittivo.
2. Usa i widget Markdown: prima di utilizzare le metriche, usa i widget Markdown per aggiungere un contesto testuale nella parte superiore della tua dashboard. Questo contesto specifica cosa include la dashboard, qual è l'importanza delle metriche rappresentate e può contenere anche link ad altre dashboard e strumenti di risoluzione dei problemi.
3. Crea le variabili della dashboard: integra le variabili della dashboard, se necessario, in modo da offrire visualizzazioni dinamiche e flessibili della dashboard.
4. Crea i widget per le metriche: aggiungi i widget per le metriche in modo da visualizzare varie metriche emesse dall'applicazione e personalizza questi widget in modo che rappresentino efficacemente lo stato del sistema e i risultati aziendali.
5. Query di Log Insights: utilizza CloudWatchLog Insights per ricavare metriche utilizzabili dai tuoi log e visualizzare queste informazioni sulla tua dashboard.
6. Configura gli allarmi: integra gli CloudWatchallarmi nella dashboard per una rapida visualizzazione di tutte le metriche che superano le relative soglie.
7. Usa Contributor Insights: incorpora CloudWatchContributor Insights per analizzare i campi ad alta cardinalità e comprendere meglio i principali contributori della tua risorsa.
8. Progetta widget personalizzati: per esigenze specifiche non soddisfatte dai widget standard, prendi in considerazione la creazione di widget personalizzati, che possono attingere da varie origini dati o rappresentare i dati in modi unici.
9. Utilizzo AWS Health Dashboard: Utilizzalo AWS Health Dashboardper ottenere informazioni più approfondite sullo stato del tuo account, sugli eventi e sulle modifiche imminenti che potrebbero influire sui tuoi servizi e risorse. Puoi anche ottenere una visualizzazione centralizzata degli eventi di integrità in AWS Organizations o creare dashboard personalizzate (per maggiori dettagli, consulta Esempi correlati).
10. Itera e perfeziona: man mano che la tua applicazione si evolve, riesamina regolarmente la dashboard per assicurarne la pertinenza.

Risorse

Best practice correlate:

- OPS04-BP01 Identifica gli indicatori chiave di prestazione
- OPS08-BP01 Analizza le metriche del carico di lavoro
- OPS08-BP02 Analizza i registri dei carichi di lavoro
- OPS08-BP03 Analizza le tracce del carico di lavoro
- OPS08-BP04 Crea avvisi utilizzabili

Documenti correlati:

- Creazione di pannelli di controllo per visibilità operativa
- Utilizzo di Amazon CloudWatch Dashboards

Video correlati:

- Crea dashboard per più account e più regioni CloudWatch
- AWS re:Invent 2021 - Gain enterprise visibility with Cloud AWS operation dashboards)

Esempi correlati:

- One Observability Workshop
- Monitoraggio delle applicazioni con Amazon CloudWatch
- AWS Health Pannelli di controllo e approfondimenti sull'intelligence degli eventi
- Visualize AWS Health events using Amazon Managed Grafana

OPS9. Come fai a comprendere lo stato delle operazioni?

Definisci, acquisisci e analizza i parametri delle operazioni per ottenere visibilità sugli eventi delle operazioni, in modo da intraprendere le azioni appropriate.

Best practice

- OPS09-BP01 Misura gli obiettivi operativi e con le metriche KPIs
- OPS09-BP02 Comunicare lo stato e le tendenze per garantire la visibilità delle operazioni

- OPS09-BP03 Rivedi le metriche operative e dai priorità al miglioramento

OPS09-BP01 Misura gli obiettivi operativi e con le metriche KPIs

Fissate alla vostra organizzazione degli obiettivi KPIs che definiscano il successo delle operazioni e stabilite che le metriche li riflettano. Definisci previsioni da utilizzare come riferimento e rivalutale regolarmente. Sviluppa meccanismi per raccogliere queste metriche dai team per la valutazione.

Risultato desiderato:

- Gli obiettivi e KPIs i team operativi dell'organizzazione sono stati pubblicati e condivisi.
- KPIsVengono stabilite metriche che li riflettono. Gli esempi possono includere:
  - Lunghezza della coda dei ticket o età media del ticket
  - Numero di ticket raggruppati per tipo di problema
  - Tempo impiegato a lavorare Problemi con o senza una procedura operativa standardizzata () SOP
  - Tempo impiegato per il ripristino dopo un push di codice non riuscito
  - Volume delle chiamate

Anti-pattern comuni:

- Le scadenze di implementazione non vengono rispettate perché gli sviluppatori sono costretti a dedicarsi alle attività di risoluzione dei problemi. I team di sviluppo chiedono più personale, ma non possono quantificarne il numero perché il tempo impiegato non può essere misurato.
- È stato installato un desk di livello 1 per gestire le chiamate degli utenti. Nel corso del tempo, sono aumentati i carichi di lavoro ma non il personale assegnato al desk di livello 1. La soddisfazione dei clienti ne risente a causa dell'aumento dei tempi di chiamata e di quelli per arrivare a una soluzione, ma la dirigenza non vede indicatori di questo problema e non intraprende azioni.
- Un carico di lavoro problematico è stato affidato a un team operativo separato per la gestione. A differenza di altri carichi di lavoro, questo non è accompagnato dalla documentazione e dai runbook adeguati. Pertanto, i team dedicano più tempo alla risoluzione dei problemi e alla gestione degli errori. Tuttavia, non esistono metriche che lo documentino, il che rende difficile comprendere le responsabilità.

Vantaggi dell'adozione di questa best practice: quando il monitoraggio del carico di lavoro mostra lo stato delle nostre applicazioni e servizi, i team operativi dedicati al monitoraggio forniscono ai proprietari informazioni dettagliate sui cambiamenti avvenuti tra i consumatori di tali carichi di lavoro, come le mutate esigenze aziendali. Misura l'efficacia di questi team e valutali rispetto agli obiettivi aziendali creando metriche in grado di riflettere lo stato delle operazioni. Le metriche possono evidenziare problemi relativi al supporto o identificare quando si verificano deviazioni rispetto a un obiettivo di livello di servizio.

Livello di rischio associato se questa best practice non fosse adottata: medio

Guida all'implementazione

Fissa un appuntamento per parlare con i leader aziendali e le parti interessate per stabilire gli obiettivi generali del servizio. Stabilisci quali devono essere i compiti dei vari team operativi e quali sfide potrebbero affrontare. Usando questi, raccogli gli indicatori chiave di prestazione (KPIs) che potrebbero riflettere questi obiettivi operativi. Questi potrebbero essere la soddisfazione del cliente, il tempo trascorso dall'ideazione della funzionalità alla sua implementazione, il tempo medio di risoluzione dei problemi e altro.

Partendo da questa KPIs base, identifica le metriche e le fonti di dati che potrebbero rispecchiare al meglio questi obiettivi. La soddisfazione del cliente può essere una combinazione di diverse metriche, come i tempi di attesa o di risposta durante le chiamate, i punteggi di soddisfazione e i tipi di problemi sollevati. I tempi di implementazione possono essere la somma del tempo necessario per il test e l'implementazione, con l'aggiunta di eventuali correzioni post-implementazione. Le statistiche che mostrano il tempo dedicato a diversi tipi di problemi (o il numero di tali problemi) possono fornire indicazioni su dove è necessario un impegno mirato.

Risorse

Documenti correlati:

- Amazon QuickSight - Utilizzo KPIs
- Amazon CloudWatch - Utilizzo delle metriche
- Creazione di pannelli di controllo
- Come monitorare l'ottimizzazione dei costi KPIs con KPI Dashboard

OPS09-BP02 Comunicare lo stato e le tendenze per garantire la visibilità delle operazioni

Conoscere lo stato delle operazioni e la direzione verso la quale tendono a muoversi è necessario per identificare quando i risultati possono essere a rischio, se è possibile supportare o meno carichi di lavoro aggiuntivi o per verificare gli effetti che le modifiche hanno avuto sui team. Durante gli eventi operativi, disporre di pagine di stato a cui gli utenti e i team operativi possono fare riferimento per ottenere informazioni può ridurre la pressione sui canali di comunicazione e diffondere informazioni in modo proattivo.

Risultato desiderato:

- I responsabili delle operazioni hanno a disposizione informazioni dettagliate per conoscere il volume di chiamate che i loro team stanno gestendo e quali operazioni sono in corso, ad esempio le implementazioni.
- Quando si verificano eventi che possono compromettere le normali operazioni, vengono inviati avvisi alle parti interessate e alle comunità di utenti.
- Quando ricevono un avviso o si verifica un problema, la leadership dell'organizzazione e le parti interessate possono controllare una pagina di stato e ottenere informazioni relative a un evento operativo, come punti di contatto, informazioni sui ticket e tempi di ripristino stimati.
- I report messi a disposizione della leadership e delle parti interessate contengono statistiche operative come il volume delle chiamate in un periodo di tempo, i punteggi di soddisfazione degli utenti, il numero e l'età di ticket in sospeso.

Anti-pattern comuni:

- Se un carico di lavoro si interrompe, il servizio diventa non disponibile. Il volume delle chiamate aumenta quando gli utenti chiedono di sapere cosa sta succedendo. Le richieste dei manager di sapere chi sta risolvendo un problema comportano un ulteriore aumento del volume. Vari team operativi duplicano gli sforzi mentre effettuano indagini.
- La volontà di acquisire una nuova capacità porta a riassegnare gli sforzi di alcuni membri del personale verso compiti di tipo tecnico. Non viene fornito alcun backfill e i tempi di risoluzione dei problemi aumentano. Queste informazioni non vengono acquisite e i manager vengono a conoscenza del problema solo dopo diverse settimane o quando viene ricevuto il feedback negativo degli utenti.

Vantaggi dell'adozione di questa best practice: a volte, durante eventi operativi che hanno un impatto sull'azienda, si spreca molto tempo ed energia in query per ottenere informazioni da vari team

nel tentativo di comprendere la situazione. Grazie alla creazione di pagine di stato e dashboard ampiamente diffuse, le parti interessate possono ottenere rapidamente informazioni, ad esempio, se è stato rilevato o meno un problema, chi è a capo delle attività di risoluzione o quando è previsto un ritorno alle normali operazioni. Ciò permette ai membri del team di avere più tempo per affrontare i problemi, perché non devono dilungarsi a comunicare lo stato agli altri.

Inoltre, pannelli di controllo e report forniscono informazioni ai responsabili delle decisioni e alle parti interessate in modo da scoprire se i team operativi sono in grado di rispondere alle esigenze aziendali e le modalità di allocazione delle relative risorse. Questo aspetto è fondamentale per determinare la presenza di risorse adeguate a supporto dell'azienda.

Livello di rischio associato se questa best practice non fosse adottata: medio

Guida all'implementazione

Crea pannelli di controllo che mostrino le metriche fondamentali attuali per i tuoi team operativi e rendile facilmente accessibili ai responsabili operativi e ai manager.

Crea pagine di stato aggiornabili rapidamente per diffondere informazioni relative a un incidente o un evento, come chi ne è responsabile e chi coordina la risposta. Condividi in questa pagina eventuali passaggi o soluzioni alternative che gli utenti dovrebbero prendere in considerazione e divulga ampiamente la posizione della pagina. Incoraggia gli utenti a controllare prima questa pagina quando si trovano di fronte a un problema sconosciuto.

Raccogli e fornisci report che mostrino le condizioni delle operazioni nel tempo e distribuiscili a leader e responsabili decisionali per illustrare il lavoro dei team operativi e le loro sfide ed esigenze.

Condividi tra i team le metriche e i report che meglio rispecchiano gli obiettivi KPIs e i punti in cui sono stati influenti nel guidare il cambiamento. Dedica del tempo a queste attività per aumentare l'importanza delle operazioni nei e tra i team.

Risorse

Documenti correlati:

- Measure Progress
- Creazione di pannelli di controllo per visibilità operativa

Soluzioni correlate:

- Data Operations

OPS09-BP03 Rivedi le metriche operative e dai priorità al miglioramento

L'assegnazione di tempo e risorse dedicati alla revisione dello stato delle operazioni garantisce che servire il day-to-day settore di attività rimanga una priorità. Effettua regolarmente riunioni con i responsabili operativi e le parti interessate per rivedere le metriche, riconfermare o modificare traguardi e obiettivi e dare priorità ai miglioramenti.

Risultato desiderato:

- I responsabili operativi e il personale si incontrano regolarmente per esaminare le metriche in un determinato periodo di riferimento. Si comunicano le sfide, si celebrano le vittorie e si condividono le lezioni apprese.
- Le parti interessate e i dirigenti aziendali vengono regolarmente informati sullo stato delle operazioni e sollecitati a fornire input sugli obiettivi KPIs e sulle iniziative future. Vengono discusse e contestualizzate le scelte tra erogazione dei servizi, operazioni e manutenzione.

Anti-pattern comuni:

- Viene lanciato un nuovo prodotto, ma i team operativi di livello 1 e 2 non sono adeguatamente formati per fornire supporto oppure non dispongono di personale aggiuntivo. I leader non vedono le metriche che mostrano la diminuzione dei tempi di risoluzione dei ticket e l'aumento del volume degli incidenti. Si agisce settimane dopo, quando i numeri delle sottoscrizioni iniziano a diminuire a causa di utenti scontenti che abbandonano la piattaforma.
- Da molto tempo esiste un processo manuale per eseguire la manutenzione su un carico di lavoro. La volontà di automatizzare, seppur presente, costituiva una priorità bassa data la scarsa importanza del sistema. Nel corso del tempo, tuttavia, l'importanza del sistema è cresciuta e ora i team operativi sono impegnati per la maggior parte del tempo in questi processi manuali. Non sono previste risorse per fornire una maggiore strumentazione ai team operativi oberati dall'aumento dei carichi di lavoro, con rischi di burnout per il personale. La leadership viene a conoscenza del problema una volta segnalato da un membro del personale che lascia l'azienda per un concorrente.

Vantaggi dell'adozione di questa best practice: in alcune organizzazioni, può diventare difficile dedicare lo stesso tempo e la stessa attenzione alla fornitura di servizi e a nuovi prodotti od offerte. Quando ciò si verifica, il settore d'attività può risentirne a causa del lento deterioramento del livello di servizio atteso. Questo perché le operazioni non cambiano e non si evolvono di pari passo con

la crescita del business e possono diventare presto obsolete. Senza una revisione regolare delle informazioni raccolte dai team operativi, il rischio che l'azienda corre potrebbe diventare visibile solo quando è troppo tardi. Dedicare tempo alla revisione delle metriche e delle procedure insieme al personale operativo e alla leadership, permette di mettere in luce il ruolo cruciale svolto dai team operativi nell'identificare i rischi molto prima che raggiungano livelli critici. I team operativi ottengono una visione migliore dei cambiamenti e delle iniziative aziendali imminenti, il che permette di intraprendere azioni proattive. Grazie alla visibilità delle metriche operative, la leadership è consapevole del ruolo che i team operativi svolgono nel garantire la soddisfazione dei clienti, sia interni che esterni, ed è in grado di valutare meglio le scelte in base alle priorità o di garantire che ci sia sufficiente tempo per modificare e fare evolvere operazioni e risorse attraverso nuove iniziative aziendali e di carico di lavoro.

Livello di rischio associato se questa best practice non fosse adottata: medio

Guida all'implementazione

Dedica del tempo alla revisione delle metriche operative con le parti interessate e i team operativi e alla revisione dei dati dei report. Inserisci questi report nel contesto degli scopi e degli obiettivi dell'organizzazione per stabilire se vengono raggiunti. Individua le cause di ambiguità in caso di obiettivi non chiari o potenziali conflitti tra quanto richiesto e quanto offerto.

Identifica come il tempo, le persone e gli strumenti possono contribuire agli esiti delle operazioni. Determina quale KPIs impatto avrebbe e quali dovrebbero essere gli obiettivi di successo. Effettua regolarmente una revisione per assicurarti che i team operativi dispongano di risorse sufficienti per supportare il settore d'attività.

Risorse

Documenti correlati:

- Amazon Athena
- Riferimento alle CloudWatch metriche e alle dimensioni di Amazon
- Amazon QuickSight
- AWS Glue
- AWS Glue Data Catalog
- Raccogli parametri e log EC2 dalle istanze Amazon e dai server locali con Amazon Agent CloudWatch

- Utilizzo dei CloudWatch parametri di Amazon

OPS10. In che modo gestisci gli eventi del carico di lavoro e delle operazioni?

Prepara e convalida le procedure in risposta agli eventi per ridurre al minimo il loro impatto sul tuo carico di lavoro.

Best practice

- OPS10-BP01 Utilizzare un processo per la gestione di eventi, incidenti e problemi
- OPS10-BP02 Avere un processo per avviso
- OPS10-BP03 Assegna priorità agli eventi operativi in base all'impatto aziendale
- OPS10-BP04 Definire i percorsi di escalation
- OPS10-BP05 Definire un piano di comunicazione con i clienti per gli eventi che hanno un impatto sui servizi
- OPS10-BP06 Comunicazione dello stato tramite dashboard
- OPS10-BP07 Automatizza le risposte agli eventi

OPS10-BP01 Utilizzare un processo per la gestione di eventi, incidenti e problemi

La capacità di gestire in modo efficiente eventi, incidenti e problemi è fondamentale per mantenere l'integrità e le prestazioni del carico di lavoro. È essenziale riconoscere e comprendere le differenze tra questi elementi per sviluppare una strategia di risposta e risoluzione efficace. Stabilire e seguire un processo ben definito per ogni aspetto facilita la gestione rapida ed efficace da parte del tuo team di qualsiasi sfida operativa che si presenti.

Risultato desiderato: la tua organizzazione gestisce efficacemente eventi operativi, incidenti e problemi attraverso processi ben documentati e archiviati a livello centrale. Questi processi vengono costantemente aggiornati per riflettere le modifiche, semplificando la gestione e mantenendo l'affidabilità del servizio e delle prestazioni dei carichi di lavoro elevata.

Anti-pattern comuni:

- Rispondi in modo reattivo, anziché proattivo, agli eventi.
- Vengono adottati approcci incoerenti a diversi tipi di eventi o incidenti.
- La tua organizzazione non effettua analisi e non impara dagli incidenti per prevenire eventi futuri.

Vantaggi dell'adozione di questa best practice:

- Processi di risposta semplificati e standardizzati.
- Riduzione dell'impatto degli incidenti su servizi e clienti.
- Risoluzione rapida dei problemi.
- Miglioramento continuo dei processi operativi.

Livello di rischio associato se questa best practice non fosse adottata: elevato

Guida all'implementazione

L'implementazione di questa best practice prevede la registrazione degli eventi dei carichi di lavoro. Per la gestione di incidenti e problemi, è necessario ricorrere ai processi. I processi sono documentati, condivisi e aggiornati con frequenza. I problemi vengono identificati, classificati in base alla priorità e corretti.

Informazioni su eventi, incidenti e problemi

- Eventi: un evento è l'adempimento di un'azione, un'occorrenza o un cambiamento di stato. Gli eventi possono essere pianificati o non pianificati e possono avere origine all'interno o all'esterno del carico di lavoro.
- Incidenti: gli incidenti sono eventi che richiedono una risposta, come interruzioni non pianificate o il peggioramento della qualità del servizio. Rappresentano interruzioni che richiedono un'attenzione immediata al fine di ripristinare il normale funzionamento del carico di lavoro.
- Problemi: i problemi sono le cause alla base di uno o più incidenti. Identificare e risolvere i problemi implica approfondire gli incidenti per prevenire eventi futuri.

Passaggi dell'implementazione

Eventi

1. Monitora gli eventi:
   - Implementa l'osservabilità e sfrutta l'osservabilità del carico di lavoro.
   - Le azioni di monitoraggio intraprese da un utente, un ruolo o un AWS servizio vengono registrate come eventi in. AWS CloudTrail
   - Rispondi ai cambiamenti operativi delle tue applicazioni in tempo reale con Amazon EventBridge.

- Valuta, monitora e registra continuamente le modifiche alla configurazione delle risorse con AWS Config.

2. Crea processi:
   - Sviluppa un processo per valutare quali eventi sono significativi e richiedono di essere monitorati. Ciò comporta l'impostazione di soglie e parametri per le attività normali e anomale.
   - Determina i criteri in base ai quali un evento viene segnalato come un incidente, ad esempio, la gravità dell'evento, l'impatto sugli utenti o la deviazione dal comportamento previsto.
   - Rivedi regolarmente i processi di monitoraggio e risposta agli eventi. Ciò include l'analisi degli incidenti passati, l'adeguamento delle soglie e il perfezionamento dei meccanismi di avviso.

Incidenti

1. Rispondi agli incidenti:
   - Usa gli approfondimenti degli strumenti di osservabilità per identificare e rispondere rapidamente agli incidenti.
   - Implementa AWS Systems Manager Ops Center per aggregare, organizzare e dare priorità agli elementi operativi e agli incidenti.
   - Utilizza servizi come Amazon CloudWatch e AWS X-Rayper analisi e risoluzione dei problemi più approfondite.
   - Prendi in considerazione AWS Managed Services (AMS) per una migliore gestione degli incidenti, sfruttando le sue capacità proattive, preventive e investigative. AMSestende il supporto operativo con servizi come il monitoraggio, il rilevamento e la risposta agli incidenti e la gestione della sicurezza.
   - Per i clienti del supporto Enterprise, AWS Incident Detection and Response offre un monitoraggio proattivo continuo e la gestione degli incidenti per i carichi di lavoro di produzione.
2. Crea un processo di gestione degli incidenti:
   - Definisci un processo strutturato di gestione degli incidenti, che includa ruoli, protocolli di comunicazione e passaggi per la risoluzione chiari.
   - Integra la gestione degli incidenti con strumenti come AWS Chatbot per garantire l'efficienza nella risposta e nel coordinamento.
   - Suddividi in categorie gli incidenti in base alla gravità, con piani di risposta agli incidenti predefiniti per ciascuna di esse.
3. Apprendi e migliora:

- Effettua analisi post-incidente per comprendere le cause principali e l'efficacia della risoluzione.
- Aggiorna e migliora continuamente i piani di risposta in base alle revisioni e alle pratiche in evoluzione.
- Documenta e condividi le lezioni apprese tra i team per migliorare la resilienza operativa.
- I clienti del supporto Enterprise possono rivolgersi al proprio Technical Account Manager per il workshop sulla gestione degli incidenti. Questo workshop guidato consente di verificare il piano di risposta agli incidenti esistente e ti aiuta a individuare eventuali aree da migliorare.

Problemi

1. Identifica i problemi:
   - Utilizza i dati degli incidenti passati per identificare modelli ricorrenti che potrebbero indicare la presenza di problemi sistemici più profondi.
   - Sfrutta strumenti come AWS CloudTrailAmazon CloudWatch per analizzare le tendenze e scoprire i problemi sottostanti.
   - Coinvolgi team interfunzionali, ad esempio i team dediti alle operazioni, allo sviluppo e i reparti aziendali, per ottenere prospettive diverse sulle cause principali.
2. Crea un processo di gestione dei problemi:
   - Sviluppa un processo strutturato per la gestione dei problemi, concentrandoti su soluzioni a lungo termine piuttosto che su correzioni rapide.
   - Incorpora tecniche di analisi delle cause principali (RCA) per indagare e comprendere le cause alla base degli incidenti.
   - Aggiorna policy e procedure operative e l'infrastruttura in base agli esiti per prevenire il ripetersi degli incidenti.
3. Continua a migliorare:
   - Promuovi una cultura di apprendimento e miglioramento continui, incoraggiando i team a identificare e affrontare in modo proattivo i problemi potenziali.
   - Analizza e rivedi regolarmente i processi e gli strumenti di gestione dei problemi per allinearli agli scenari aziendali e tecnologici in evoluzione.
   - Condividi approfondimenti e best practice in tutta l'organizzazione per creare un ambiente operativo più resiliente ed efficiente.
4. Impegnarsi AWS Support:

- Utilizza risorse di AWS supporto, ad esempio AWS Trusted Advisorper indicazioni proattive e consigli di ottimizzazione.
- I clienti del supporto Enterprise hanno a disposizione programmi dedicati, come AWS Countdown, per ricevere assistenza durante gli eventi critici.

Livello di impegno per il piano di implementazione: medio

Risorse

Best practice correlate:

- OPS04-BP01 Identifica gli indicatori chiave di prestazione
- OPS04-BP02 Implementare la telemetria delle applicazioni
- OPS07-BP03 Usa i runbook per eseguire le procedure
- OPS07-BP04 Usa i playbook per analizzare i problemi
- OPS08-BP01 Analizza le metriche del carico di lavoro
- OPS11-BP02 Eseguire l'analisi post-incidente

Documenti correlati:

- AWS Security Incident Response Guide
- AWS Rilevamento e risposta agli incidenti
- AWS Framework di adozione del cloud: prospettiva operativa - Gestione degli incidenti e dei problemi
- La gestione degli incidenti nell'era del DevOps e SRE
- PagerDuty - Cos'è la gestione degli incidenti?

Video correlati:

- I migliori consigli di risposta agli incidenti di AWS
- AWS re:Invent 2022 - The Amazon Builders' Library: 25 anni di eccellenza operativa Amazon
- AWS re:Invent 2022 - Rilevamento e risposta agli incidenti (01) AWS SUP2
- Presentazione di Incident Manager da AWS Systems Manager

Esempi correlati:

- AWS Servizi proattivi — Workshop sulla gestione degli incidenti
- Come automatizzare la risposta agli incidenti con e PagerDuty AWS Systems Manager Incident Manager
- Coinvolgi i soccorritori agli incidenti con gli orari di chiamata in AWS Systems Manager Incident Manager
- Migliora la visibilità e la collaborazione durante la gestione degli incidenti in AWS Systems Manager Incident Manager
- Segnalazioni di incidenti e richieste di assistenza in AMS

Servizi correlati:

- Amazon EventBridge

OPS10-BP02 Avere un processo per avviso

Stabilire un processo chiaro e definito per ogni avviso nel sistema è essenziale per una gestione degli incidenti efficace ed efficiente. Questa pratica garantisce che ogni avviso porti a una risposta specifica e attuabile, migliorando l'affidabilità e la reattività delle operazioni.

Risultato desiderato: ogni avviso avvia un piano di risposta specifico e ben definito. Ove possibile, le risposte sono automatizzate e dotate di una chiara titolarità e di un percorso di escalation definito. Gli avvisi sono collegati a una base di up-to-date conoscenze in modo che qualsiasi operatore possa rispondere in modo coerente ed efficace. Le risposte sono rapide e uniformi su tutta la linea, migliorando l'efficienza e l'affidabilità operativa.

Anti-pattern comuni:

- Gli avvisi non hanno un processo di risposta predefinito, il che porta a risoluzioni improvvisate e tardive.
- Il sovraccarico di avvisi comporta che gli avvisi importanti vengano trascurati.
- Gli avvisi vengono gestiti in modo incoerente a causa della mancanza di titolarità e responsabilità chiare.

Vantaggi dell'adozione di questa best practice:

- Creazione solo di avvisi utilizzabili, con conseguente riduzione dell'affaticamento da avvisi.
- Riduzione del tempo medio di risoluzione (MTTR) dei problemi operativi.
- Riduzione del tempo medio di indagine (MTTI), che contribuisce a ridurreMTTR.
- Migliore capacità di scalare le risposte operative.
- Maggiore coerenza e affidabilità nella gestione degli eventi operativi.

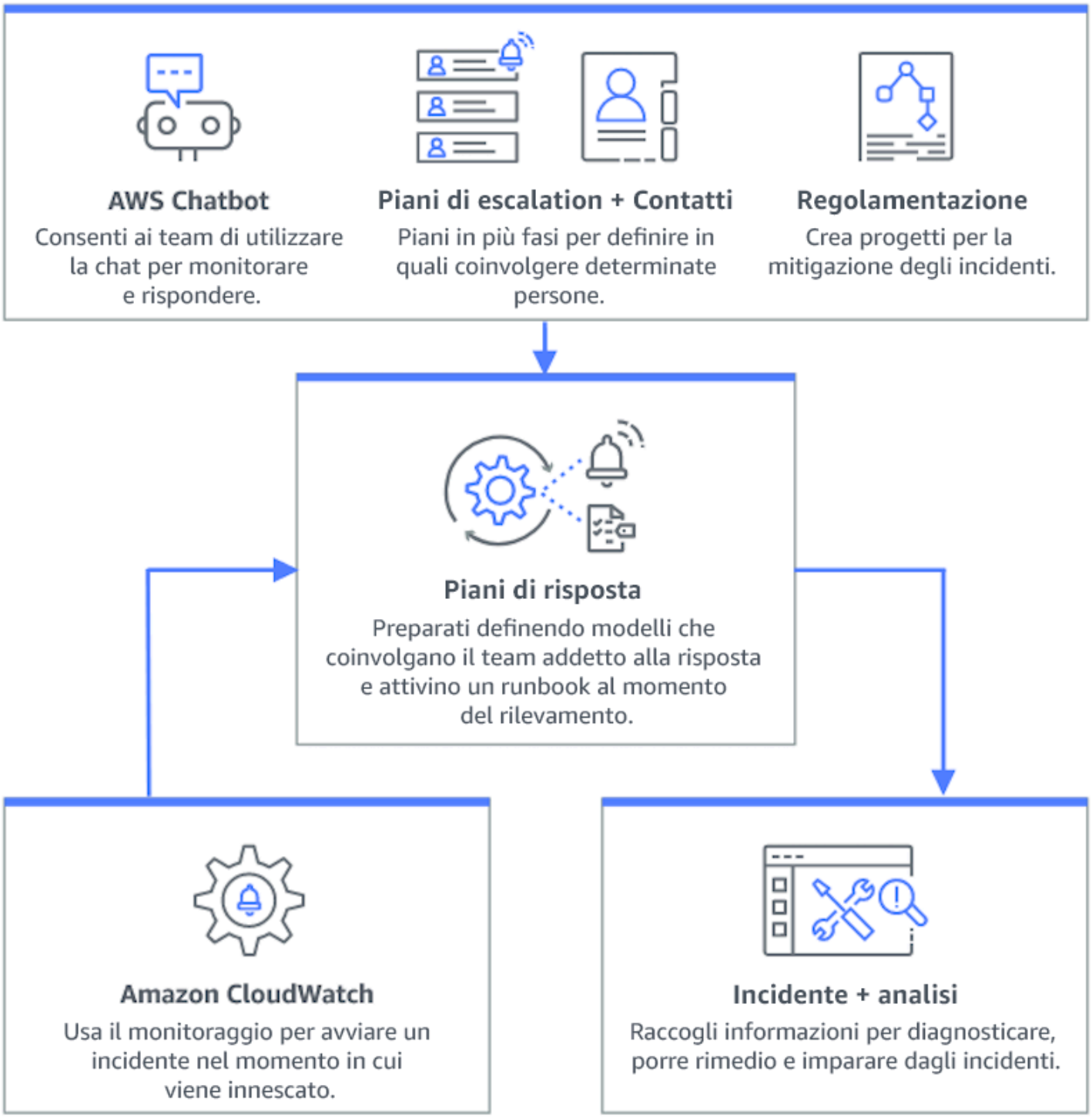
Livello di rischio associato se questa best practice non fosse adottata: elevato

Guida all'implementazione

Avere un processo per ogni avviso implica stabilire un piano di risposta chiaro per ciascun avviso, automatizzare le risposte ove possibile e perfezionare continuamente questi processi in base al feedback operativo e all'evoluzione dei requisiti.

Passaggi dell'implementazione

Il diagramma seguente illustra il flusso di lavoro di gestione degli incidenti all'interno di AWS Systems Manager Incident Manager. È progettato per rispondere rapidamente ai problemi operativi creando automaticamente incidenti in risposta a eventi specifici di Amazon o CloudWatch Amazon. EventBridge Quando viene creato un incidente, automaticamente o manualmente, Incident Manager centralizza la gestione dell'incidente, organizza le informazioni pertinenti sulle AWS risorse e avvia piani di risposta predefiniti. Ciò include l'esecuzione dei runbook di Systems Manager Automation per un'azione immediata e la creazione di un elemento di lavoro operativo principale OpsCenter per tenere traccia delle attività e delle analisi correlate. Questo processo semplificato velocizza e coordina la risposta agli incidenti in tutto l'ambiente. AWS



1. Usa allarmi compositi: crea allarmi compositi CloudWatch per raggruppare allarmi correlati, riducendo il rumore e consentendo risposte più significative.
2. Integra gli CloudWatch allarmi di Amazon con Incident Manager Configura gli CloudWatch allarmi per creare automaticamente incidenti in. AWS Systems Manager Incident Manager

3. Integra Amazon EventBridge con Incident Manager: crea EventBridge regole per reagire agli eventi e crea incidenti utilizzando piani di risposta definiti.
4. Preparati per gli incidenti in Incident Manager:
    - Crea piani di risposta dettagliati in Incident Manager per ciascun tipo di avviso.
    - Stabilisci canali di chat tramite AWS Chatbot collegato ai piani di risposta in Incident Manager, semplificando la comunicazione in tempo reale durante gli incidenti su piattaforme come Slack, Microsoft Teams e Amazon Chime.
    - Integra i runbook di Systems Manager Automation in Incident Manager per fornire risposte automatiche agli incidenti.

Risorse

Best practice correlate:

- OPS04-BP01 Identifica gli indicatori chiave di prestazione
- OPS08-BP04 Crea avvisi utilizzabili

Documenti correlati:

- AWS Cloud Adoption Framework: Prospettiva operativa - Gestione degli incidenti e dei problemi
- Utilizzo degli CloudWatch allarmi Amazon
- Configurazione AWS Systems Manager Incident Manager
- Preparing for incidents in Incident Manager

Video correlati:

- I migliori consigli di risposta agli incidenti di AWS

Esempi correlati:

- AWS Workshop - AWS Systems Manager Incident Manager - Automatizza la risposta agli incidenti agli eventi di sicurezza

OPS10-BP03 Assegna priorità agli eventi operativi in base all'impatto aziendale

Rispondere tempestivamente agli eventi operativi è fondamentale, ma non tutti gli eventi sono uguali. Quando si assegnano le priorità in base all'impatto sul business, si dà la priorità anche alla risoluzione di eventi che possono avere conseguenze significative, come la compromissione della sicurezza, perdite finanziarie, violazioni normative o danni alla reputazione.

Risultato desiderato: la priorità delle risposte agli eventi operativi si basa sul potenziale impatto dell'evento su operazioni e obiettivi di business. Ciò rende le risposte efficienti ed efficaci.

Anti-pattern comuni:

- Ogni evento viene trattato con lo stesso livello di urgenza, generando confusione e ritardi nell'affrontare le criticità.
- Non è possibile distinguere tra eventi ad alto e basso impatto, con conseguente errata allocazione delle risorse.
- L'organizzazione non dispone di un chiaro framework di assegnazione delle priorità, il che genera risposte incoerenti agli eventi operativi.
- Agli eventi viene assegnata la priorità in base all'ordine in cui vengono segnalati piuttosto che al loro impatto sui risultati aziendali.

Vantaggi dell'adozione di questa best practice:

- Assicura che la risposta si concentri in primo luogo sulle funzioni aziendali critiche, riducendo al minimo i danni potenziali.
- Migliora l'allocazione delle risorse durante più eventi simultanei.
- Migliora la capacità dell'organizzazione di mantenere la fiducia e soddisfare i requisiti normativi.

Livello di rischio associato se questa best practice non fosse adottata: elevato

Guida all'implementazione

Di fronte a molteplici eventi operativi, è essenziale un approccio strutturato alla definizione delle priorità basato sull'impatto e sull'urgenza. Questo approccio consente di prendere decisioni informate, indirizzare gli sforzi laddove sono più necessari e mitigare il rischio per la continuità aziendale.

Passaggi dell'implementazione

1. Valuta l'impatto: sviluppa un sistema di classificazione per valutare la gravità degli eventi in termini di potenziale impatto sulle operazioni e sugli obiettivi di business. L'esempio seguente mostra le categorie di impatto:

Livello di impatto	Descrizione
Elevata	Coinvolge molti dipendenti o clienti, ha un elevato impatto finanziario, genera un elevato danno alla reputazione o lesioni.
Media	Coinvolge un gruppo di dipendenti o clienti, ha un impatto finanziario moderato o genera un danno alla reputazione moderato.
Bassa	Coinvolge singoli dipendenti o clienti, ha un basso impatto finanziario o genera un danno alla reputazione di lieve entità.

2. Valuta l'urgenza: definisci i livelli di urgenza in base alla rapidità con cui un evento richiede una risposta, considerando fattori quali sicurezza, implicazioni finanziarie e accordi sui livelli di servizio (). SLAs L'esempio seguente illustra le categorie di urgenza:

Livello di urgenza	Descrizione
Elevata	Danni in aumento esponenziale, impatto sui lavori urgenti, aumento imminente della situazione o impatto su utenti o gruppi. VIP
Media	I danni aumentano nel tempo o vengono colpiti singoli utenti o gruppi. VIP
Bassa	I danni marginali aumentano nel tempo o influiscono non-time-sensitive sul lavoro.

3. Crea una matrice di prioritizzazione:
   - Usa una matrice per incrociare impatto e urgenza, assegnando livelli di priorità a diverse combinazioni.

- Rendi la matrice accessibile e comprensibile da tutti i membri del team responsabili delle risposte agli eventi operativi.
- La seguente matrice di esempio mostra la gravità dell'incidente in base all'urgenza e all'impatto:

Urgenza e impatto	Elevata	Media	Bassa
Elevata	Critico	Urgente	Elevata
Media	Urgente	Elevata	Normale
Bassa	Elevata	Normale	Bassa

4. Predisponi formazione e comunicazione: forma i team di risposta sulla matrice di prioritizzazione e sull'importanza di attenersi a essa durante un evento. Comunica il processo di definizione delle priorità a tutte le parti interessate per stabilire aspettative chiare.
5. Integra con la risposta agli incidenti:
   - Incorpora la matrice di prioritizzazione nei tuoi piani e strumenti di risposta agli incidenti.
   - Automatizza la classificazione e la prioritizzazione degli eventi, ove possibile, per accelerare i tempi di risposta.
   - I clienti del supporto Enterprise, possono sfruttare AWS Incident Detection and Response che garantisce il monitoraggio proattivo 24 ore su 24, 7 giorni su 7, oltre alla gestione degli incidenti per i carichi di lavoro di produzione.
6. Rivedi e adatta: rivedi regolarmente l'efficacia del processo di definizione delle priorità e apporta modifiche in base al feedback e ai cambiamenti nell'ambiente aziendale.

Risorse

Best practice correlate:

- OPS03-BP03 L'escalation è incoraggiata
- OPS08-BP04 Crea avvisi utilizzabili
- OPS09-BP01 Misura gli obiettivi operativi e con le metriche KPIs

Documenti correlati:

- Atlassian - Understanding incident severity levels

- IT Process Map - Checklist Incident Priority

OPS10-BP04 Definire i percorsi di escalation

Stabilisci percorsi di escalation chiari all'interno dei tuoi protocolli di risposta agli incidenti per facilitare un'azione tempestiva ed efficace. Ciò include la specificazione delle istruzioni per l'escalation, la descrizione dettagliata del processo di escalation e l'approvazione preventiva delle azioni per accelerare il processo decisionale e ridurre il tempo medio di risoluzione (). MTTR

Risultato desiderato: un processo strutturato ed efficiente che inoltra gli incidenti al personale appropriato, riducendo al minimo i tempi di risposta e l'impatto.

Anti-pattern comuni:

- La mancanza di chiarezza in merito alle procedure di ripristino genera risposte improvvisate in caso di incidenti critici.
- L'assenza di autorizzazioni e titolarità definite comporta ritardi quando è necessaria un'azione urgente.
- Le parti interessate e i clienti non sono informati nei tempi attesi.
- Le decisioni importanti subiscono ritardi.

Vantaggi dell'adozione di questa best practice:

- Risposta semplificata agli incidenti tramite procedure di escalation predefinite.
- Tempi di inattività ridotti con azioni preapprovate e titolarità chiara.
- Migliore allocazione delle risorse e adeguamenti del livello di supporto in base alla gravità degli incidenti.
- Migliore comunicazione con le parti interessate e i clienti.

Livello di rischio associato se questa best practice non fosse adottata: medio

Guida all'implementazione

I percorsi di escalation definiti correttamente sono fondamentali per una risposta rapida agli incidenti. AWS Systems Manager Incident Manager supporta l'impostazione di piani di escalation strutturati e pianificazioni di chiamata, che avvisano il personale giusto in modo che sia pronto ad agire in caso di incidenti.

Passaggi dell'implementazione

1. Imposta le richieste di escalation: imposta gli allarmi per creare un incidente. CloudWatch AWS Systems Manager Incident Manager
2. Imposta la pianificazione della reperibilità: crea la pianificazione della reperibilità in Incident Manager, in linea con i tuoi percorsi di escalation. Fornisci al personale di turno le autorizzazioni e gli strumenti necessari per agire rapidamente.
3. Procedure di escalation dettagliate:
   - Determina le condizioni specifiche in base alle quali un incidente deve essere inoltrato.
   - Crea piani di escalation in Incident Manager.
   - I canali di escalation devono consistere in un contatto o in una pianificazione della reperibilità.
   - Definisci i ruoli e le responsabilità del team a ogni livello di escalation.
4. Approva preventivamente le azioni di mitigazione: collabora con i responsabili delle decisioni per approvare preventivamente le azioni per gli scenari previsti. Sfrutta i runbook di Systems Manager Automation integrati con Incident Manager per velocizzare la risoluzione degli incidenti.
5. Specifica la proprietà: identifica chiaramente i proprietari interni per ogni fase del percorso di escalation.
6. Fornisci dettagli in merito alle escalation a terze parti:
   - Documenta gli accordi sui livelli di servizio di terze parti (SLAs) e allineali agli obiettivi interni.
   - Stabilisci protocolli chiari per la comunicazione con i fornitori durante gli incidenti.
   - Integra i contatti dei fornitori negli strumenti di gestione degli incidenti per l'accesso diretto.
   - Conduci regolarmente esercitazioni che includano scenari di risposta di terze parti.
   - Mantieni le informazioni sulle escalation dei fornitori ben documentate e facilmente accessibili.
7. Esegui formazione e test per i piani di escalation: forma il tuo team sul processo di escalation e conduci regolarmente esercitazioni di risposta agli incidenti o giornate di gioco. I clienti del supporto Enterprise possono richiedere un workshop sulla gestione degli incidenti.
8. Continua a migliorare: verifica regolarmente l'efficacia dei tuoi percorsi di escalation. Aggiorna i tuoi processi in base alle lezioni apprese dalle analisi degli incidenti e dal feedback continuo.

Livello di impegno per il piano di implementazione: moderato

Risorse

Best practice correlate:

- OPS08-BP04 Crea avvisi utilizzabili
- OPS10-BP02 Avere un processo per avviso
- OPS11-BP02 Eseguire l'analisi post-incidente

Documenti correlati:

- AWS Systems Manager Incident Manager Piani di escalation
- Working with on-call schedules in Incident Manager
- Creating and Managing Runbooks
- Gestione temporanea degli accessi elevati con AWS IAM Identity Center
- Atlassian - Escalation policies for effective incident management

OPS10-BP05 Definire un piano di comunicazione con i clienti per gli eventi che hanno un impatto sui servizi

Una comunicazione efficace durante gli eventi che incidono sul servizio è fondamentale per mantenere la fiducia e la trasparenza con i clienti. Un piano di comunicazione ben definito sostiene la comunicazione rapida e chiara di informazioni all'interno e all'esterno dell'organizzazione durante gli incidenti.

Risultato desiderato:

- Un solido piano di comunicazione che informa efficacemente i clienti e le parti interessate durante gli eventi che influiscono sul servizio.
- Trasparenza nella comunicazione per creare fiducia e ridurre la preoccupazione dei clienti.
- Riduzione al minimo dell'impatto che gli eventi che incidono sul servizio hanno sull'esperienza del cliente e sulle operazioni aziendali.

Anti-pattern comuni:

- Una comunicazione inadeguata o in ritardo genera confusione e insoddisfazione nei clienti.
- Una messaggistica eccessivamente tecnica o vaga impedisce la comunicazione dell'impatto effettivo sugli utenti.
- È assente una strategia di comunicazione predefinita, con conseguente messaggistica incoerente e reattiva.

Vantaggi dell'adozione di questa best practice:

- Maggiore fiducia e soddisfazione dei clienti attraverso una comunicazione chiara e proattiva.
- Riduzione del carico operativo per i team di supporto grazie alla risoluzione preventiva delle preoccupazioni dei clienti.
- Maggiore efficienza di gestione e risoluzione degli incidenti.

Livello di rischio associato se questa best practice non fosse adottata: medio

Guida all'implementazione

La creazione di un piano di comunicazione completo per gli eventi che incidono sul servizio implica prendere in considerazione molteplici aspetti, dalla scelta dei canali giusti alla creazione del messaggio e del tono. Il piano deve essere adattabile, scalabile e soddisfare diversi scenari di interruzione del servizio.

Passaggi dell'implementazione

1. Definisci ruoli e responsabilità:
   - Assegna a un responsabile degli incidenti gravi la supervisione delle attività di risposta agli incidenti.
   - Designa un responsabile delle comunicazioni dedicato al coordinamento di tutte le comunicazioni esterne e interne.
   - Includi il responsabile dell'assistenza per fornire una comunicazione coerente attraverso ticket di supporto.
2. Identifica i canali di comunicazione: seleziona canali come chat sul posto di lavoro, e-mail, social mediaSMS, notifiche in-app e pagine di stato. Questi canali devono essere resilienti e in grado di operare in maniera indipendente durante gli eventi che incidono sul servizio.
3. Comunica in modo rapido, chiaro e regolare con i clienti:
   - Sviluppa modelli per vari scenari di compromissione del servizio, focalizzandoti sulla semplicità e sui dettagli essenziali. Includi informazioni sul problema relativo al servizio, sui tempi di risoluzione previsti e sull'impatto.
   - Usa Amazon Pinpoint per avvisare i clienti tramite notifiche push, notifiche in-app, e-mail, SMS, messaggi vocali e messaggi su canali personalizzati.
   - Usa Amazon Simple Notification Service (AmazonSNS) per avvisare gli abbonati in modo programmatico o tramite e-mail, notifiche push mobili e messaggi di testo.

- Comunica lo stato tramite dashboard condividendo pubblicamente una CloudWatch dashboard di Amazon.
- Incoraggia il coinvolgimento sui social media:
  - Monitora attivamente i social media per comprendere il sentimento dei clienti.
  - Pubblica post su piattaforme di social media per aggiornare il pubblico e coinvolgere la comunità.
  - Prepara modelli per una comunicazione coerente e chiara sui social media.

4. Coordina la comunicazione interna: implementa i protocolli interni utilizzando strumenti come AWS Chatbot il coordinamento e la comunicazione del team. Usa i CloudWatch dashboard per comunicare lo stato.
5. Orchestra la comunicazione con strumenti e servizi dedicati:
   - Utilizzalo AWS Systems Manager Incident Manager con AWS Chatbot per configurare canali di chat dedicati per la comunicazione e il coordinamento interni in tempo reale durante gli incidenti.
   - Usa AWS Systems Manager Incident Manager i runbook per automatizzare le notifiche ai clienti tramite Amazon Pinpoint, SNS Amazon o strumenti di terze parti come le piattaforme di social media durante gli incidenti.
   - Incorpora i flussi di lavoro di approvazione all'interno dei runbook per rivedere e autorizzare tutte le comunicazioni esterne prima dell'invio.
6. Fai pratica e migliora:
   - Tieni corsi di formazione sull'uso di strumenti e strategie di comunicazione. Responsabilizza i team affinché siano in grado di prendere decisioni tempestive durante gli incidenti.
   - Testa il piano di comunicazione con esercitazioni regolari o giornate di gioco. Usa questi test per perfezionare la messaggistica e valutare l'efficacia dei canali.
   - Implementa meccanismi di feedback per valutare l'efficacia della comunicazione durante gli incidenti. Sviluppa continuamente il piano di comunicazione in base al feedback e alle esigenze mutevoli.

Livello di impegno per il piano di implementazione: elevato

Risorse

Best practice correlate:

- OPS07-BP03 Usa i runbook per eseguire le procedure

- OPS10-BP06 Comunicazione dello stato tramite dashboard
- OPS11-BP02 Eseguire l'analisi post-incidente

Documenti correlati:

- Atlassian - Incident communication best practices
- Atlassian - How to write a good status update
- PagerDuty - Una guida alle comunicazioni relative agli incidenti

Video correlati:

- Atlassian - Create your own incident communication plan: Incident templates

Esempi correlati:

- AWS Health Dashboard
- Esempi di aggiornamenti AWS di stato

OPS10-BP06 Comunicazione dello stato tramite dashboard

Usa i pannelli di controllo come strumento strategico per trasmettere lo stato operativo e le metriche fondamentali in tempo reale a diversi tipi di pubblico, inclusi team tecnici interni, leader e clienti. Questi pannelli di controllo offrono una rappresentazione visiva centralizzata dello stato del sistema e delle prestazioni aziendali, il che migliora la trasparenza e l'efficienza decisionale.

Risultato desiderato:

- I pannelli di controllo forniscono una visione completa del sistema e delle metriche aziendali rilevanti per le varie parti interessate.
- Le parti interessate possono accedere in modo proattivo alle informazioni operative, il che riduce la necessità di richieste di stato frequenti.
- Migliore processo decisionale in tempo reale durante le normali operazioni e gli incidenti.

Anti-pattern comuni:

- I tecnici che partecipano a una chiamata di gestione degli incidenti hanno bisogno di ricevere aggiornamenti di stato per poter agire rapidamente.
- Affidarsi ai report manuali per la gestione comporta ritardi e potenziali imprecisioni.
- I team operativi vengono spesso interrotti per aggiornamenti sullo stato durante gli incidenti.

Vantaggi dell'adozione di questa best practice:

- Consente alle parti interessate di accedere immediatamente alle informazioni critiche, promuovendo un processo decisionale informato.
- Riduce le inefficienze operative riducendo al minimo i report manuali e le richieste di stato frequenti.
- Aumenta la trasparenza e la fiducia attraverso la visibilità in tempo reale delle prestazioni del sistema e delle metriche aziendali.

Livello di rischio associato se questa best practice non fosse adottata: medio

Guida all'implementazione

I pannelli di controllo comunicano efficacemente lo stato dei sistemi e le metriche aziendali e possono essere personalizzati in base alle esigenze di diversi gruppi di destinatari. Strumenti come Amazon CloudWatch dashboard e Amazon ti QuickSight aiutano a creare dashboard interattivi in tempo reale per il monitoraggio del sistema e la business intelligence.

Passaggi dell'implementazione

1. Identifica le esigenze delle parti interessate: determina le esigenze in termini di informazioni specifiche dei diversi gruppi di destinatari, come team tecnici, leader e clienti.
2. Scegli gli strumenti giusti: seleziona gli strumenti appropriati come i CloudWatch dashboard di Amazon per il monitoraggio del sistema e Amazon QuickSight per la business intelligence interattiva.
3. Progetta pannelli di controllo efficaci:
   - Progetta dashboard per presentare chiaramente le metriche pertinenti e KPIs assicurarti che siano comprensibili e utilizzabili.
   - Incorpora visualizzazioni a livello di sistema e a livello aziendale, se necessario.
   - Includi pannelli di controllo di alto livello (per ampie panoramiche) e di basso livello (per analisi dettagliate).

- Integra allarmi automatici all'interno dei pannelli di controllo per evidenziare i problemi critici.
- Annota i pannelli di controllo con soglie e obiettivi delle metriche importanti per una visibilità immediata.

4. Integra l'origine dati:
   - Usa Amazon CloudWatch per aggregare e visualizzare i parametri di vari AWS servizi e interrogare i parametri da altre fonti di dati, creando una visione unificata delle metriche aziendali e di salute del tuo sistema.
   - Usa funzionalità come CloudWatch Logs Insights per interrogare e visualizzare i dati di log provenienti da diverse applicazioni e servizi.
5. Fornisci l'accesso self-service:
   - Condividi le CloudWatch dashboard con le parti interessate per l'accesso alle informazioni in modalità self-service utilizzando le funzionalità di condivisione delle dashboard.
   - Assicurati che le dashboard siano facilmente accessibili e forniscano informazioni in tempo reale. up-to-date
6. Aggiorna e perfeziona regolarmente:
   - Aggiorna e perfeziona continuamente i panelli di controllo per allinearli alle esigenze aziendali in evoluzione e ai feedback delle parti interessate.
   - Rivedi regolarmente i panelli di controllo per assicurarti che siano sempre pertinenti ed efficaci nella trasmissione delle informazioni necessarie.

Risorse

Best practice correlate:

- OPS08-BP05 Crea dashboard

Documenti correlati:

- Creazione di pannelli di controllo per visibilità operativa
- Utilizzo delle CloudWatch dashboard di Amazon
- Create flexible dashboards with dashboard variables
- Dashboard di condivisione CloudWatch
- Query metrics from other data sources
- Aggiungi un widget personalizzato a una dashboard CloudWatch

Esempi correlati:

- One Observability Workshop - Dashboards

OPS10-BP07 Automatizza le risposte agli eventi

L'automazione delle risposte agli eventi è fondamentale per una gestione operativa rapida, coerente e priva di errori. Crea processi semplificati e utilizza strumenti per gestire e rispondere automaticamente agli eventi, riducendo al minimo gli interventi manuali e migliorando l'efficacia operativa.

Risultato desiderato:

- Riduzione degli errori umani e tempi di risoluzione più rapidi grazie all'automazione.
- Gestione degli eventi operativi coerente e affidabile.
- Maggiore efficienza operativa e affidabilità del sistema.

Anti-pattern comuni:

- La gestione manuale degli eventi comporta ritardi ed errori.
- L'automazione viene trascurata nelle attività ripetitive e critiche.
- Le attività manuali ripetitive causano affaticamento da avvisi e la mancata identificazione di problemi critici.

Vantaggi dell'adozione di questa best practice:

- Risposte agli eventi accelerate, riduzione dei tempi di inattività del sistema.
- Operazioni affidabili con gestione automatizzata e coerente degli eventi.

Livello di rischio associato se questa best practice non fosse adottata: medio

Guida all'implementazione

Incorpora l'automazione per creare flussi di lavoro operativi efficienti e ridurre al minimo gli interventi manuali.

Passaggi dell'implementazione

1. Identifica le opportunità di automazione: definisci le attività ripetitive da automatizzare, come la risoluzione dei problemi, l'arricchimento dei ticket, la gestione della capacità, la scalabilità, le implementazioni e i test.
2. Identifica i prompt di automazione:
   - Valuta e definisci condizioni o metriche specifiche che avviano risposte automatiche utilizzando le azioni di CloudWatch allarme di Amazon.
   - Usa Amazon EventBridge per rispondere agli eventi nei AWS servizi, nei carichi di lavoro personalizzati e nelle applicazioni SaaS.
   - Prendi in considerazione eventi di avvio come voci di registro specifiche, soglie di metriche prestazionali o cambiamenti di stato nelle risorse. AWS
3. Implementa l'automazione basata sugli eventi:
   - Utilizza i runbook di AWS Systems Manager automazione per semplificare le attività di manutenzione, implementazione e correzione.
   - La creazione di incidenti in Incident Manager raccoglie e aggiunge automaticamente dettagli sulle AWS risorse coinvolte nell'incidente.
   - Monitora in modo proattivo le quote utilizzando Quota Monitor for AWS.
   - Regola in automatico la capacità di AWS Auto Scaling così da mantenere disponibilità e prestazioni.
   - Automatizza le pipeline di sviluppo con Amazon. CodeCatalyst
   - Smoke testa o monitora continuamente gli endpoint utilizzando il monitoraggio sintetico. APIs
4. Esegui la mitigazione del rischio attraverso l'automazione:
   - Implementa le risposte di sicurezza automatizzate per affrontare in modo rapido i rischi.
   - Utilizza AWS Systems Manager State Manager per ridurre la deviazione delle configurazioni.
   - Risolvi le risorse non conformi con. Regole di AWS Config

Livello di impegno per il piano di implementazione: elevato

Risorse

Best practice correlate:

- OPS08-BP04 Crea avvisi utilizzabili
- OPS10-BP02 Avere un processo per avviso

Documenti correlati:

- Using Systems Manager Automation runbooks with Incident Manager
- Creating incidents in Incident Manager
- AWS quote di servizio
- Monitor resource usage and send notifications when approaching quotas
- AWS Auto Scaling
- Che cos'è Amazon CodeCatalyst?
- Utilizzo degli CloudWatch allarmi Amazon
- Utilizzo delle azioni di CloudWatch allarme di Amazon
- Correzione delle risorse non conformi con Regole di AWS Config
- Creating metrics from log events using filters
- AWS Systems Manager State Manager

Video correlati:

- Crea runbook di automazione con AWS Systems Manager
- Come automatizzare le operazioni IT su AWS
- AWS Security Hub regole di automazione
- Avvia rapidamente il tuo progetto software con CodeCatalyst i blueprints di Amazon

Esempi correlati:

- CodeCatalyst Tutorial Amazon: creazione di un progetto con il modello di applicazione Web moderno a tre livelli
- One Observability Workshop
- Respond to incidents using Incident Manager

Evoluzione

Domanda

- OPS11. In che modo fai evolvere le operazioni?

OPS11. In che modo fai evolvere le operazioni?

Dedica tempo e risorse per ottenere un miglioramento incrementale pressoché continuo, per far evolvere l'efficacia e l'efficienza delle tue operazioni.

Best practice

- OPS11-BP01 Avere un processo per il miglioramento continuo
- OPS11-BP02 Eseguire l'analisi post-incidente
- OPS11-BP03 Implementazione di circuiti di feedback
- OPS11-BP04 Eseguire la gestione della conoscenza
- OPS11-BP05 Definire i fattori di miglioramento
- OPS11-BP06 Validare gli approfondimenti
- OPS11-BP07 Revisioni delle metriche di Perform operations
- OPS11-BP08 Documenta e condividi le lezioni apprese
- OPS11-BP09 Dedica tempo per apportare miglioramenti

OPS11-BP01 Avere un processo per il miglioramento continuo

Valuta il carico di lavoro rispetto alle best practice dell'architettura interna ed esterna. Effettua revisioni frequenti e deliberate del carico di lavoro. Dai priorità alle opportunità di miglioramento nella cadenza di sviluppo del software.

Risultato desiderato:

- Analizza di frequente il carico di lavoro rispetto alle best practice dell'architettura.
- Stabilisci per le opportunità di miglioramento la stessa priorità che assegni alle funzionalità del processo di sviluppo software.

Anti-pattern comuni:

- Non hai condotto una revisione dell'architettura del carico di lavoro da quando è stato implementato diversi anni fa.
- Assegni una priorità inferiore alle opportunità di miglioramento. Rispetto alle nuove funzionalità, queste opportunità rimangono nel backlog.
- Non esiste uno standard per l'implementazione delle modifiche alle best practice per l'organizzazione.

Vantaggi dell'adozione di questa best practice:

- Il carico di lavoro si basa up-to-date sulle migliori pratiche di architettura.
- Fai evolvere il carico di lavoro in modo intenzionale.
- Puoi utilizzare le best practice dell'organizzazione per migliorare tutti i carichi di lavoro.
- Ottieni guadagni marginali che hanno un impatto cumulativo, con un incremento dell'efficienza.

Livello di rischio associato se questa best practice non fosse adottata: elevato

Guida all'implementazione

Effettui di frequente la revisione dell'architettura del carico di lavoro. Utilizzi le best practice interne ed esterne per valutare il carico di lavoro e identificare le opportunità di miglioramento. Dai priorità alle opportunità di miglioramento nella cadenza di sviluppo del software.

Passaggi dell'implementazione

1. Esegui la revisione periodica dell'architettura del carico di lavoro di produzione secondo una frequenza concordata. Utilizza uno standard architettonico documentato che includa best AWS practice specifiche.
   a. Usa gli standard definiti internamente per queste revisioni. Se non disponi di standard interni, usa il Framework AWS Well-Architected.
   b. Utilizzatelo AWS Well-Architected Tool per creare una panoramica personalizzata delle vostre best practice interne e condurre una revisione dell'architettura.
   c. Contatta il tuo AWS Solution Architect o il Technical Account Manager per condurre una revisione guidata di Well-Architected Framework del tuo carico di lavoro.
2. Dai priorità alle opportunità di miglioramento identificate durante la revisione nel processo di sviluppo del software.

Livello di impegno per il piano di implementazione: basso Puoi utilizzare AWS Well-Architected Framework per condurre la tua revisione annuale dell'architettura.

Risorse

Best practice correlate:

- OPS11-BP02 Esegue l'analisi post-incidente
- OPS11-BP08 Documenta e condividi le lezioni apprese

- OPS04 Implementa l'osservabilità

Documenti correlati:

- AWS Well-Architected Tool - Lenti personalizzate
- Whitepaper AWS Well-Architected: il processo di revisione
- Personalizza le recensioni Well-Architected utilizzando obiettivi personalizzati e AWS Well-Architected Tool
- Implementazione del AWS ciclo di vita delle lenti personalizzate Well-Architected nella tua organizzazione

Video correlati:

- Well-Architected Labs - Level 100: lenti personalizzate AWS Well-Architected Tool
- AWS re:Invent 2023 - Scalabilità delle migliori pratiche Well-Architected AWS in tutta l'organizzazione

Esempi correlati:

- AWS Well-Architected Tool

OPS11-BP02 Eseguire l'analisi post-incidente

Esamina gli eventi che influiscono sui clienti e identifica i fattori che contribuiscono e le azioni preventive. Utilizza queste informazioni per sviluppare modi per limitare o prevenire il ripetersi degli incidenti. Sviluppa procedure per attivare risposte rapide ed efficaci. Comunica i fattori che hanno contribuito al presentarsi dell'imprevisto e le azioni correttive secondo necessità, specificamente mirate per il pubblico di destinazione.

Risultato desiderato:

- Stabilisci processi di gestione degli incidenti che includono l'analisi post-incidente.
- Hai a disposizione piani di osservabilità per raccogliere dati sugli eventi.
- Con questi dati comprendi e raccogli metriche che supportano il tuo processo di analisi post-incidente.
- Impari dagli incidenti per migliorare i risultati futuri.

Anti-pattern comuni:

- Sei amministratore di un server di applicazioni. Circa ogni 23 ore e 55 minuti tutte le sessioni attive vengono terminate. Hai tentato di identificare ciò che non va a buon fine sul server di applicazioni. Sospetti che potrebbe trattarsi di un problema di rete, ma non riesci a ottenere la collaborazione dal team di rete perché i suoi membri sono troppo occupati per supportarti. Ti manca un processo predefinito da seguire per ottenere supporto e raccogliere le informazioni necessarie per stabilire che cosa sta accadendo.
- Si è verificata una perdita di dati all'interno del carico di lavoro. Questa è la prima volta che si è verificata e la causa non è immediatamente identificabile. Decidi che non è importante perché puoi ricreare i dati. La perdita di dati inizia a verificarsi con maggiore frequenza e influisce sui clienti. Questo comporta inoltre un ulteriore onere operativo quando ripristini i dati mancanti.

Vantaggi dell'adozione di questa best practice:

- Disponendo di un processo predefinito per determinare i componenti, le condizioni, le azioni e gli eventi che hanno contribuito a un incidente, sei in grado di identificare le opportunità di miglioramento.
- Utilizzi i dati dell'analisi post-incidente per apportare miglioramenti.

Livello di rischio associato se questa best practice non fosse adottata: elevato

Guida all'implementazione

Utilizza un processo per determinare i fattori determinanti. Esamina tutti gli incidenti che influiscono sul cliente. Predisponi un processo per identificare e documentare i fattori che contribuiscono a un incidente, in modo da sviluppare azioni di mitigazione in grado di limitare o impedire il suo ripetersi e per sviluppare procedure che consentano risposte rapide ed efficaci. Comunica le cause principali degli incidenti in modo appropriato e personalizza la comunicazione in base al pubblico di destinazione. Condividi quanto appreso in maniera aperta all'interno della tua organizzazione.

Passaggi dell'implementazione

1. Raccogli metriche come le modifiche all'implementazione e alla configurazione, l'ora di inizio dell'incidente, l'ora dell'allarme, dell'intervento, dell'inizio della mitigazione e il tempo di risoluzione dell'incidente.
2. Descrivi i momenti fondamentali sulla linea temporale per comprendere gli eventi dell'incidente.

3. Poniti le seguenti domande:
    a. Potresti migliorare il tempo di rilevamento?
    b. Sono presenti aggiornamenti alle metriche e agli allarmi che permettono di rilevare l'incidente prima?
    c. Puoi migliorare i tempi di diagnosi?
    d. Sono presenti aggiornamenti ai tuoi piani di risposta o di escalation che potrebbero coinvolgere prima i team di risposta corretti?
    e. Puoi migliorare il tempo necessario per la mitigazione?
    f. Ci sono passaggi del runbook o del playbook che potresti aggiungere o migliorare?
    g. È possibile prevenire che si verifichino incidenti futuri?
4. Crea liste di controllo e azioni. Monitora ed esegui tutte le azioni.

Livello di impegno per il piano di implementazione: medio

Risorse

Best practice correlate:

- OPS11-BP01 Avere un processo per il miglioramento continuo
- OPS4 - Implementare l'osservabilità

Documenti correlati:

- Performing a post-incident analysis in Incident Manager
- Revisione della prontezza operativa

OPS11-BP03 Implementazione di circuiti di feedback

I cicli di feedback forniscono informazioni fruibili che guidano il processo decisionale. Vanno creati nelle procedure e nei carichi di lavoro per identificare i problemi e le aree che necessitano di miglioramenti. Inoltre, convalidano gli investimenti effettuati nei miglioramenti. Questi cicli di feedback sono la base per migliorare continuamente il carico di lavoro.

Sono due le categorie dei cicli di feedback: feedback immediato e analisi retrospettiva. Il feedback immediato viene raccolto con la revisione delle prestazioni e dei risultati delle attività operative. Questo feedback proviene dai membri del team, dai clienti o dall'output automatizzato dell'attività. Il

feedback immediato viene ricevuto ad esempio dal test A/B e dall'offerta di nuove funzionalità, ed è essenziale per anticipare l'errore (fail fast).

L'analisi retrospettiva viene eseguita regolarmente per acquisire il feedback della revisione dei risultati operativi e dei parametri nel tempo. Queste retrospettive si svolgono alla fine di uno sprint, in base a una cadenza o dopo importanti rilasci o eventi. Questo tipo di ciclo di feedback convalida gli investimenti nelle operazioni o nel carico di lavoro, consente di misurare il successo e comprova la tua strategia.

Risultato desiderato: utilizzi feedback immediato e analisi retrospettiva per apportare miglioramenti. L'applicazione di un meccanismo per acquisire il feedback di utenti e membri del team. L'uso dell'analisi retrospettiva per identificare le tendenze che guidano i miglioramenti.

Anti-pattern comuni:

- Lanci una nuova funzionalità ma non hai modo di ricevere il feedback dei clienti.
- Dopo aver investito in miglioramenti delle operazioni, non conduci una retrospettiva per convalidare gli investimenti.
- Raccogli il feedback dei clienti ma non lo esamini regolarmente.
- I cicli di feedback portano alla proposta di elementi di azione non sono inclusi nel processo di sviluppo software.
- I clienti non ricevono un feedback sui miglioramenti che hanno proposto.

Vantaggi dell'adozione di questa best practice:

- Puoi lavorare a ritroso con il cliente per promuovere nuove funzionalità.
- La cultura della tua organizzazione può reagire più rapidamente ai cambiamenti.
- Le tendenze vengono utilizzate per identificare le opportunità di miglioramento.
- Le retrospettive convalidano gli investimenti effettuati per il carico di lavoro e le operazioni.

Livello di rischio associato se questa best practice non fosse adottata: elevato

Guida all'implementazione

L'implementazione di questa best practice comporta l'utilizzo del feedback immediato e dell'analisi retrospettiva. Questi cicli di feedback promuovono i miglioramenti. Esistono molti meccanismi per il feedback immediato, inclusi questionari, sondaggi dei clienti o moduli di feedback. La tua

organizzazione utilizza anche le retrospettive per identificare le opportunità di miglioramento e convalidare le iniziative.

Esempio del cliente

AnyCompany Retail ha creato un modulo web in cui i clienti possono fornire feedback o segnalare problemi. Durante lo scrum settimanale, il feedback degli utenti viene valutato dal team di sviluppo software. Il feedback viene regolarmente utilizzato per guidare l'evoluzione della piattaforma. Viene eseguita una retrospettiva alla fine di ogni sprint per identificare gli elementi che devono essere migliorati.

Passaggi dell'implementazione

1. Feedback immediato
   - Hai bisogno di un meccanismo per ricevere il feedback dai clienti e dai membri del team. Le attività operative possono anche essere configurate per fornire un feedback automatizzato.
   - L'organizzazione ha bisogno di un processo per rivedere il feedback, determinare cosa migliorare e pianificare il miglioramento.
   - Il feedback deve essere aggiunto al processo di sviluppo software.
   - Quando apporti miglioramenti, contatta l'autore del feedback.
     - Puoi utilizzarlo AWS Systems Manager OpsCenterper creare e tenere traccia di questi miglioramenti come OpsItems.
2. Analisi retrospettiva
   - Conduci le retrospettive alla fine di un ciclo di sviluppo, a una cadenza prestabilita o dopo un rilascio importante.
   - Riunisci le parti interessate coinvolte nel carico di lavoro per la riunione retrospettiva.
   - Crea tre colonne sulla lavagna o in un foglio di lavoro: Fine, Inizio e Mantenimento.
     - Fine riguarda per tutto ciò che vuoi che il team smetta di fare.
     - Inizio riguarda per le idee che vuoi iniziare ad applicare.
     - Mantenimento indica ciò che vuoi continuare a fare.
   - Raccogli il feedback dalle parti interessate.
   - Dai priorità al feedback. Assegna le azioni e le parti interessate a qualsiasi elemento nelle colonne Inizio e Mantenimento.
   - Aggiungi le azioni al processo di sviluppo software e comunica gli aggiornamenti sullo stato alle parti interessate mentre apporti i miglioramenti.

Livello di impegno per il piano di implementazione: medio Per implementare questa best practice è necessario un modo per ricevere il feedback immediato e analizzarlo. Inoltre, è necessario stabilire un processo di analisi retrospettiva.

Risorse

Best practice correlate:

- OPS01-BP01 Valuta le esigenze dei clienti: i cicli di feedback sono un meccanismo per raccogliere le esigenze dei clienti esterni.
- OPS01-BP02 Valuta le esigenze interne dei clienti: le parti interessate interne possono utilizzare i cicli di feedback per comunicare necessità e requisiti.
- OPS11-BP02 Eseguire l'analisi post-incidente: le analisi successive agli incidenti sono una forma importante di analisi retrospettiva da condurre dopo gli incidenti.
- OPS11-BP07 Revisioni delle metriche di Perform operations: le revisioni dei parametri operativi identificano tendenze e aree di miglioramento.

Documenti correlati:

- 7 insidie da evitare quando si costruisce un CCOE
- Atlassian Team Playbook - Retrospectives
- Email Definitions: Feedback Loops
- Stabilire cicli di feedback basati sulla revisione del AWS Well-Architected Framework
- IBMGarage Methodology - Organizza una retrospettiva
- Investopedia — Il ciclo PDCS
- Maximizing Developer Effectiveness by Tim Cochran
- White paper su Operations Readiness Reviews (ORR) - Iterazione
- ITILCSI- Miglioramento continuo del servizio
- When Toyota met e-commerce: Lean at Amazon

Video correlati:

- Building Effective Customer Feedback Loops

Esempi correlati:

- Astuto - Open source customer feedback tool
- AWS Soluzioni - Q on nABot AWS
- Fider: una piattaforma per organizzare il feedback dei clienti)

Servizi correlati:

- AWS Systems Manager OpsCenter

OPS11-BP04 Eseguire la gestione della conoscenza

La gestione delle informazioni permette ai membri del team di trovare le informazioni necessarie per svolgere il proprio lavoro. Nelle organizzazioni che promuovono la formazione dei propri dipendenti, le informazioni vengono liberamente condivise, migliorando le competenze personali. Le informazioni possono essere vagliate o cercate. Le informazioni sono accurate e aggiornate. Esistono meccanismi per creare nuove informazioni, aggiornare quelle esistenti e archiviare quelle obsolete. L'esempio più comune di una piattaforma di gestione delle informazioni è un sistema di gestione dei contenuti come un wiki.

Risultato desiderato:

- Accesso per i membri del team a informazioni tempestive e accurate.
- Possibilità di eseguire ricerche nelle informazioni.
- Presenza di un meccanismo per aggiungere, aggiornare e archiviare le informazioni.

Anti-pattern comuni:

- Assenza di un sistema di archiviazione centrale delle informazioni. I membri del team gestiscono i propri appunti su computer locali.
- Presenza di un wiki self-hosted, ma senza alcun meccanismo per la gestione delle informazioni, con informazioni non aggiornate di conseguenza.
- Le informazioni mancanti vengono identificate da qualcuno, ma non esiste un processo per richiederne l'aggiunta nel wiki del team. I dipendenti le aggiungono manualmente ma omettono un passaggio importante, causando un'interruzione.

Vantaggi dell'adozione di questa best practice:

- I membri del team acquisiscono le competenze necessarie perché le informazioni vengono condivise liberamente.
- Nuovi membri del team vengono integrati più facilmente perché la documentazione è aggiornata e può essere oggetto di ricerche.
- Le informazioni sono tempestive, accurate e di utilità pratica.

Livello di rischio associato se questa best practice non fosse adottata: elevato

Guida all'implementazione

La gestione delle informazioni è un aspetto importante delle aziende che promuovono la formazione dei propri dipendenti. Per iniziare, è necessario un repository centrale in cui archiviare le informazioni, un esempio comune del quale è un wiki self-hosted. Devi sviluppare processi per l'aggiunta, l'aggiornamento e l'archiviazione delle informazioni. Sviluppa standard per gli aspetti da documentare e permetti a ciascuno di contribuire.

Esempio del cliente

AnyCompany Retail ospita un Wiki interno in cui sono archiviate tutte le conoscenze. I membri del team sono incoraggiati ad aggiungere il proprio input nella knowledge base durante lo svolgimento delle proprie mansioni quotidiane. Ogni trimestre un team interfunzionale valuta le pagine obsolete e determina se devono essere archiviate o aggiornate.

Passaggi dell'implementazione

1. Per iniziare, identifica il sistema di gestione dei contenuti in cui verranno archiviate le informazioni. Ottieni il consenso delle parti interessate in tutta l'organizzazione.
    a. Se non possiedi un sistema di gestione dei contenuti, valuta se affidarti a un wiki self-hosted o usare un repository con controllo delle versioni come punto di partenza.
2. Sviluppa runbook per l'aggiunta, l'aggiornamento e l'archiviazione delle informazioni. Fornisci ai team la formazione necessaria su questi processi.
3. Identifica le informazioni che devono essere archiviate nel sistema di gestione dei contenuti. Inizia dalle attività quotidiane (runbook e playbook) svolte dai membri del team. Collabora con le parti interessate per classificare in ordine di priorità le informazioni aggiunte.
4. Collabora periodicamente con le parti interessate per identificare out-of-date le informazioni e archiviarle o aggiornarle.

Livello di impegno per il piano di implementazione: medio Se non possiedi un sistema di gestione dei contenuti, puoi configurare un wiki self-hosted o un repository di documenti con controllo delle versioni.

Risorse

Best practice correlate:

- OPS11-BP08 Documenta e condividi le lezioni apprese: la gestione delle informazioni semplifica la condivisione delle conclusioni sulle lezioni apprese.

Documenti correlati:

- Atlassian - Knowledge Management

Esempi correlati:

- DokuWiki
- Gollum
- MediaWiki
- Wiki.js

OPS11-BP05 Definire i fattori di miglioramento

Identifica i fattori che promuovono il miglioramento in modo da valutare e dare priorità alle opportunità sulla base di dati e cicli di feedback. Esplora le opportunità di miglioramento nei sistemi e nei processi e automatizza laddove appropriato.

Risultato desiderato:

- Tieni traccia dei dati provenienti da tutto l'ambiente.
- Esegui la correlazione di eventi e attività ai risultati aziendali.
- Puoi confrontare e contrapporre ambienti e sistemi.
- Mantieni una cronologia dettagliata delle attività relative alle implementazioni e ai risultati.
- Raccogli i dati a supporto del livello di sicurezza.

Anti-pattern comuni:

- Raccogli dati da tutto l'ambiente, ma non correli eventi e attività.
- Raccogli dati dettagliati da tutto il tuo patrimonio e ciò favorisce un aumento di Amazon, AWS CloudTrail attività CloudWatch e costi. tuttavia non utilizzi questi dati in modo significativo.
- Non tieni conto dei risultati aziendali quando definisci i fattori che promuovono il miglioramento.
- Non misuri gli effetti delle nuove funzionalità.

Vantaggi dell'adozione di questa best practice:

- Determinando i criteri di miglioramento, riduci al minimo l'impatto delle motivazioni basate sugli eventi o degli investimenti influenzati da fattori emotivi.
- Rispondi agli eventi aziendali, non solo a quelli tecnici.
- Misuri l'ambiente per identificare le aree di miglioramento.

Livello di rischio associato se questa best practice non fosse adottata: medio

Guida all'implementazione

- Comprensione dei fattori che promuovono il miglioramento: è consigliabile apportare modifiche a un sistema solo quando un risultato desiderato è supportato.
  - Funzionalità desiderate: prendi in considerazione le funzionalità e le capacità desiderate quando valuti le opportunità di miglioramento.
    - Cosa c'è di nuovo con AWS
  - Problemi inaccettabili: tieni in considerazione i problemi, i bug e le vulnerabilità inaccettabili quando valuti le opportunità di miglioramento. Tieni traccia delle giuste opzioni di ridimensionamento corretto e individua le opportunità di ottimizzazione.
    - Bollettini sulla sicurezza AWS aggiornati
    - AWS Trusted Advisor
    - Cloud Intelligence Dashboards
  - Requisiti di conformità: quando esamini le opportunità di miglioramento, prendi in considerazione gli aggiornamenti e le modifiche necessarie per mantenere la conformità a normative e policy o per avere diritto al supporto di terze parti.
    - Conformità di AWS
    - Programmi per la conformità di AWS
    - Ultime novità sulla conformità di AWS

Risorse

Best practice correlate:

- OPS01 Priorità organizzative
- OPS02 Relazioni e proprietà
- OPS04-BP01 Identifica gli indicatori chiave di prestazione
- OPS08 Utilizzo dell'osservabilità del carico di lavoro
- OPS09 Comprendere l'Operational Health
- OPS11-BP03 Implementa cicli di feedback

Documenti correlati:

- Amazon Athena
- Amazon QuickSight
- Conformità di AWS
- Ultime novità sulla conformità di AWS
- Programmi per la conformità di AWS
- AWS Glue
- Bollettini sulla sicurezza AWS aggiornati
- AWS Trusted Advisor
- Export your log data to Amazon S3
- Novità di AWS
- Gli aspetti imprescindibili dell'innovazione orientata al cliente
- Digital Transformation: Hype or a Strategic Necessity?

Video correlati

- AWS re:Invent 2023 - Migliora l'efficienza operativa e la resilienza con (0) AWS Support SUP31

OPS11-BP06 Validare gli approfondimenti

Rivedi i risultati dell'analisi e le risposte con i team trasversali e i proprietari dell'azienda. Utilizza queste revisioni per definire una visione comune, identificare ulteriori impatti e stabilire le linee d'azione. Adatta le risposte, se necessario.

Risultati desiderati:

- Rivedi regolarmente gli approfondimenti con i proprietari dell'azienda. Gli imprenditori forniscono un contesto aggiuntivo alle informazioni appena acquisite.
- Esamini gli approfondimenti e richiedi il feedback ai colleghi tecnici, quindi condividi le tue conoscenze con i team.
- Pubblichi i dati e gli approfondimenti affinché altri team tecnici e aziendali possano esaminarli. Tieni conto di quanto appreso nelle nuove procedure di altri reparti.
- Riassumi ed esamini i nuovi approfondimenti con i leader senior. I leader senior utilizzano i nuovi approfondimenti per definire la strategia.

Anti-pattern comuni:

- Rilasci una nuova funzionalità che modifica alcuni comportamenti dei clienti. La tua osservabilità non tiene conto di queste modifiche. Non quantifichi i vantaggi di queste modifiche.
- Invii un nuovo aggiornamento e trascuri di aggiornare il tuo. CDN La CDN cache non è più compatibile con l'ultima versione. Misuri la percentuale di richieste con errori. Tutti i tuoi utenti segnalano HTTP 400 errori durante la comunicazione con i server di backend. Analizzi gli errori del cliente e scopri che, avendo misurato la dimensione sbagliata, il tuo tempo è stato improduttivo.
- L'accordo sul livello di servizio prevede un tempo di attività del 99,9% e l'obiettivo del punto di ripristino è di quattro ore. Il proprietario del servizio sostiene che il sistema non subisce tempi di inattività. Implementi una soluzione di replica costosa e complessa, che comporta uno spreco di tempo e denaro.

Vantaggi dell'adozione di questa best practice:

- Convalidando gli approfondimenti con i proprietari dell'azienda e con gli esperti in materia, è possibile stabilire una comprensione comune e gestire il miglioramento in modo più efficace.
- Individui i problemi nascosti e ne tieni conto nelle decisioni future.
- La tua attenzione passa dai risultati tecnici ai risultati aziendali.

Livello di rischio associato se questa best practice non fosse adottata: medio

Guida all'implementazione

- Convalida delle informazioni: interagisci con i responsabili aziendali e gli esperti in materia per garantire la comprensione e l'accordo comuni sul significato dei dati raccolti. Individua ulteriori problemi e impatti potenziali e stabilisci le azioni da intraprendere.

Risorse

Best practice correlate:

- OPS01-BP06 Valuta i compromessi gestendo vantaggi e rischi
- OPS02-BP06 Le responsabilità tra i team sono predefinite o negoziate
- OPS11-BP03 Implementa cicli di feedback

Documenti correlati:

- Progettazione di un centro di eccellenza cloud () CCOE

Video correlati:

- Building observability to increase resiliency

OPS11-BP07 Revisioni delle metriche di Perform operations

Esegui regolarmente un'analisi retrospettiva dei parametri operativi con i partecipanti di vari team da diverse aree dell'azienda. Utilizza queste revisioni per identificare opportunità di miglioramento e potenziali linee d'azione e per condividere le conoscenze acquisite. Cerca opportunità di miglioramento in tutti i tuoi ambienti, ad esempio sviluppo, test e produzione.

Risultato desiderato:

- Esamini di frequente le metriche che hanno un impatto sull'azienda.
- Rilevi ed esamini le anomalie con le tue capacità di osservabilità.
- Utilizzi i dati per supportare i risultati e gli obiettivi aziendali.

Anti-pattern comuni:

- La finestra di manutenzione interrompe un'importante promozione al dettaglio. L'azienda non è al corrente del fatto che i normali interventi di manutenzione possono essere rimandati nel caso vi siano altri eventi di particolare rilievo per l'azienda.
- A causa dell'uso comune di una libreria obsoleta nella tua organizzazione, si è verificata una prolungata interruzione del servizio. In seguito, hai eseguito la migrazione a una libreria supportata. Gli altri team della tua organizzazione non sanno di essere a rischio.
- Non controllate regolarmente i risultati raggiunti dai clienti. SLAs Avete la tendenza a non soddisfare i vostri clienti. SLAs Sono previste sanzioni pecuniarie legate al mancato rispetto del cliente. SLAs

Vantaggi dell'adozione di questa best practice:

- Durante le riunioni che organizzi regolarmente per esaminare le metriche operative, gli eventi e gli incidenti, stabilisci una comprensione comune tra i team.
- Il tuo team si riunisce regolarmente per esaminare metriche e incidenti, il che ti consente di agire sui rischi e riconoscere i clienti. SLAs
- Condividi le lezioni apprese, che forniscono dati per la definizione delle priorità e miglioramenti mirati per ottenere i risultati aziendali.

Livello di rischio associato se questa best practice non fosse adottata: medio

Guida all'implementazione

- Esegui regolarmente un'analisi retrospettiva dei parametri operativi con i partecipanti di vari team da diverse aree dell'azienda.
- Coinvolgi le parti interessate, compresi i team che si occupano di business, sviluppo e operazioni, per convalidare gli esiti del feedback immediato e dall'analisi retrospettiva e per condividere le conoscenze acquisite.
- Utilizza gli approfondimenti di cui dispongono per identificare opportunità di miglioramento e possibili linee d'azione.

Risorse

Best practice correlate:

- OPS08-BP05 Crea dashboard

- OPS09-BP03 Rivedi le metriche operative e dai priorità al miglioramento
- OPS10-BP01 Utilizza un processo per la gestione di eventi, incidenti e problemi

Documenti correlati:

- Amazon CloudWatch
- Riferimento alle CloudWatch metriche e alle dimensioni di Amazon
- Publish custom metrics
- Utilizzo dei CloudWatch parametri di Amazon
- Dashboard e visualizzazioni con CloudWatch

OPS11-BP08 Documenta e condividi le lezioni apprese

Documenta e condividi le conoscenze acquisite durante le attività operative per metterle a frutto internamente e nei vari team. La condivisione di quanto appreso dai team comporta maggiori vantaggi all'interno dell'organizzazione. Condividi informazioni e risorse per impedire che si verifichino errori evitabili e semplificare le attività di sviluppo e concentrati sulla distribuzione delle funzionalità desiderate.

Usa AWS Identity and Access Management (IAM) per definire le autorizzazioni che consentono l'accesso controllato alle risorse che desideri condividere all'interno e tra gli account.

Risultato desiderato:

- Utilizzi repository dotati di controllo delle versioni per condividere librerie dell'applicazione, procedure di scripting, documentazione di procedure e altra documentazione di sistema.
- Condividi gli standard dell'infrastruttura come modelli AWS CloudFormation con controllo delle versioni.
- Riesamini le lezioni apprese con i team.

Anti-pattern comuni:

- Per l'uso comune di una libreria contenente degli errori nella tua organizzazione si è verificata una prolungata interruzione del servizio. Successivamente hai eseguito la migrazione a una libreria affidabile. Gli altri team della tua organizzazione non sanno di essere a rischio. Nessuno documenta e condivide l'esperienza relativa a questa libreria e nessuno è consapevole del rischio.

- Hai identificato un caso limite in un microservizio condiviso internamente che causa l'interruzione delle sessioni. Hai aggiornato le chiamate al servizio per evitare questo caso limite. Gli altri team della tua organizzazione non sanno di essere a rischio.
- Hai trovato un modo per ridurre in modo significativo i requisiti di CPU utilizzo di uno dei tuoi microservizi. Non sai se altri team potrebbero sfruttare questa tecnica.

Vantaggi dell'adozione di questa best practice: condividi le lezioni apprese a supporto del miglioramento e per trarre il massimo vantaggio dall'esperienza.

Livello di rischio associato se questa best practice non fosse adottata: basso

Guida all'implementazione

- Documenta e condividi le conoscenze acquisite: predisponi procedure per documentare le conoscenze acquisite dall'esecuzione delle attività operative e dalle analisi retrospettive affinché tali informazioni possano essere utilizzate dal altri team.
- Condividi le conoscenze acquisite: predisponi procedure per condividere con tutti i team le conoscenze acquisite e gli artefatti associati. Ad esempio condividi le procedure, le istruzioni, la governance e le best practice aggiornate tramite un wiki accessibile. Condividi script, codice e librerie tramite un repository comune.
  - Delegare l'accesso al proprio ambiente AWS
  - Condividi un repository AWS CodeCommit

Risorse

Best practice correlate:

- OPS02-BP06 Le responsabilità tra i team sono predefinite o negoziate
- OPS05-BP01 Usa il controllo della versione
- OPS05-BP06 Condividi gli standard di progettazione
- OPS11-BP03 Implementa cicli di feedback
- OPS11-BP07 Esegui revisioni delle metriche operative

Documenti correlati:

- Riduci i ritardi nei progetti con una soluzione docs-as-code

Video correlati:

- Delegare l'accesso al tuo ambiente AWS
- AWS Support s You | Exploring the Incident Management Tabletop Exercise

OPS11-BP09 Dedica tempo per apportare miglioramenti

Dedica tempo e risorse all'interno dei processi per rendere possibile il miglioramento incrementale continuo.

Risultato desiderato:

- Crei duplicati temporanei paralleli di ambienti per ridurre il rischio, lo sforzo e il costo della sperimentazione e dell'esecuzione di test.
- Questi ambienti duplicati possono essere utilizzati per testare le conclusioni di analisi ed esperimenti, ma anche per sviluppare e testare i miglioramenti pianificati.
- Gestisci gamedays e utilizzi Fault Injection Service (FIS) per fornire i controlli e i guardrail necessari ai team per eseguire esperimenti in un ambiente simile alla produzione.

Anti-pattern comuni:

- Si è verificato un problema di prestazioni noto nel server di applicazioni. Il problema viene aggiunto al backlog, dopo l'implementazione prevista delle varie funzionalità. Se la velocità con cui vengono aggiunte le funzionalità pianificate rimane costante, il problema di prestazioni non verrà mai risolto.
- Per supportare il miglioramento continuo, autorizzi amministratori e sviluppatori a utilizzare tutto il loro tempo aggiuntivo per definire e implementare miglioramenti. I miglioramenti non vengono mai completati.
- L'accettazione operativa è stata completata e non si testano più le procedure operative.

Vantaggi dell'adozione di questa best practice: dedicando tempo e risorse all'interno dei processi, puoi rendere possibile il miglioramento incrementale continuo.

Livello di rischio associato se questa best practice non fosse adottata: basso

Guida all'implementazione

- Allocazione di tempo per apportare miglioramenti: dedica tempo e risorse all'interno dei processi per rendere possibili miglioramenti graduali e continui.

- Implementa modifiche per migliorare e valutare i risultati per favorire il successo.
- Se i risultati non sono in linea con gli obiettivi e il miglioramento resta prioritario, valuta procedure d'azione alternative.
- Simula i carichi di lavoro di produzione durante le giornate di gioco e utilizza le conoscenze conseguite da queste simulazioni per migliorare.

Risorse

Best practice correlate:

- OPS05-BP08 Usa più ambienti

Video correlati:

- AWS re:Invent 2023 - Migliora la resilienza delle applicazioni con il servizio Fault Injection AWS

Sicurezza

Il pilastro della sicurezza contempla la capacità di proteggere dati, sistemi e asset per sfruttare le tecnologie cloud in modo da migliorare la sicurezza. Puoi trovare linee guida prescrittive sull'implementazione nel whitepaper sul pilastro della sicurezza.

Aree delle best practice

- Nozioni di base sulla sicurezza
- Gestione dell'identità e degli accessi
- Rilevamento
- Protezione dell'infrastruttura
- Protezione dei dati
- Risposta agli incidenti
- Sicurezza delle applicazioni

Nozioni di base sulla sicurezza

Domanda

SEC1. Come gestire un carico di lavoro in sicurezza?

Per gestire il carico di lavoro in modo sicuro, è necessario applicare le best practice globali a ogni area di sicurezza. Segui i requisiti e i processi definiti in termini di eccellenza operativa a livello organizzativo e di carico di lavoro e applicali a tutte le aree. Rimanere aggiornati con le raccomandazioni del settore AWS e l'intelligence sulle minacce vi aiuta a far evolvere il vostro modello di minaccia e gli obiettivi di controllo. L'automazione dei processi di sicurezza, i test e la convalida permettono di dimensionare le operazioni di sicurezza.

Best practice

- SEC01-BP01 Separare i carichi di lavoro utilizzando gli account
- SEC01-BP02 Utente root e proprietà dell'account sicuro
- SEC01-BP03 Identificare e convalidare gli obiettivi di controllo
- SEC01-BP04 Rimani aggiornato sulle minacce alla sicurezza e sui consigli
- SEC01-BP05 Ridurre l'ambito di gestione della sicurezza
- SEC01-BP06 Automatizza l'implementazione dei controlli di sicurezza standard
- SEC01-BP07 Identifica le minacce e dai priorità alle mitigazioni utilizzando un modello di minaccia
- SEC01-BP08 Valuta e implementa regolarmente nuovi servizi e funzionalità di sicurezza

SEC01-BP01 Separare i carichi di lavoro utilizzando gli account

Definisci guardrail e isolamento comuni tra ambienti (ad esempio, quelli di produzione, sviluppo e test) e carichi di lavoro mediante una strategia multi-account. La separazione a livello di account è fortemente consigliata, in quanto fornisce un solido confine di isolamento in termini di sicurezza, fatturazione e accesso.

Risultato desiderato: una struttura di account in grado di isolare operazioni cloud, carichi di lavoro non correlati e ambienti in account separati, così da aumentare la sicurezza nell'infrastruttura cloud.

Anti-pattern comuni:

- Inserimento di più carichi di lavoro non correlati con diversi livelli di sensibilità dei dati nello stesso account.

- Scarsa definizione della struttura dell'unità organizzativa (UO).

Vantaggi dell'adozione di questa best practice:

- Riduzione dell'impatto in caso di accesso involontario a un carico di lavoro.
- Governance centrale dell'accesso ai AWS servizi, alle risorse e alle regioni.
- Garanzia di sicurezza dell'infrastruttura cloud grazie a policy e amministrazione centralizzata dei servizi di sicurezza.
- Processo automatizzato di creazione e mantenimento dell'account.
- Audit centralizzati della tua infrastruttura per la conformità e i requisiti normativi.

Livello di rischio associato se questa best practice non fosse adottata: elevato

Guida all'implementazione

Account AWS forniscono un limite di isolamento di sicurezza tra carichi di lavoro o risorse che operano a diversi livelli di sensibilità. AWS fornisce strumenti per gestire i carichi di lavoro cloud su larga scala attraverso una strategia multi-account per sfruttare questo limite di isolamento. Per indicazioni sui concetti, i modelli e l'implementazione di una strategia multi-account su AWS, consulta Organizzazione dell'ambiente utilizzando più account AWS.

Account AWS In caso di gestione centralizzata di più account, è necessario organizzare gli account in una gerarchia definita da livelli di unità organizzative ()OUs. I controlli di sicurezza possono quindi essere organizzati e applicati agli account OUs e ai membri, stabilendo controlli preventivi coerenti sugli account dei membri dell'organizzazione. I controlli di sicurezza sono ereditati e consentono di filtrare le autorizzazioni disponibili per gli account membri situati ai livelli inferiori di una gerarchia di unità organizzative. Un buon progetto sfrutta questa ereditarietà per ridurre il numero e la complessità delle policy di sicurezza necessarie per raggiungere i controlli desiderati per ciascun account membro.

AWS Organizationse AWS Control Towersono due servizi che è possibile utilizzare per implementare e gestire questa struttura multi-account nel proprio AWS ambiente. AWS Organizations consente di organizzare gli account in una gerarchia definita da uno o più livelli diOUs, con ogni unità organizzativa contenente un numero di account membri. Le politiche di controllo del servizio (SCPs) consentono all'amministratore dell'organizzazione di stabilire controlli preventivi granulari sugli account dei membri e AWS Configpossono essere utilizzate per stabilire controlli proattivi e investigativi sugli account dei membri. Molti AWS servizi si integrano AWS Organizations per fornire

controlli amministrativi delegati ed eseguire attività specifiche del servizio su tutti gli account dei membri dell'organizzazione.

Inoltre AWS Organizations, AWS Control Towerfornisce una configurazione delle migliori pratiche con un clic per un AWS ambiente multi-account con una landing zone. La zona di destinazione è il punto di ingresso nell'ambiente multi-account stabilito da Control Tower. Control Tower offre diversi vantaggi rispetto a AWS Organizations. Tre sono i vantaggi che consentono di migliorare la governance degli account:

- Controlli di sicurezza obbligatori integrati applicati in automatico agli account ammessi nell'organizzazione.
- Controlli opzionali che possono essere attivati o disattivati per un determinato set di. OUs
- AWS Control Tower Account Factory fornisce la distribuzione automatizzata di account contenenti linee di base e opzioni di configurazione preapprovate all'interno dell'organizzazione.

Passaggi dell'implementazione

1. Progettazione di una struttura delle unità organizzative: una struttura delle unità organizzative progettata in modo corretto riduce l'onere di gestione richiesto per creare e mantenere policy di controllo dei servizi e altri controlli di sicurezza. La struttura delle unità organizzative deve essere allineata a esigenze aziendali, sensibilità dei dati e struttura del carico di lavoro.
2. Creazione di una zona di destinazione per il tuo ambiente multi-account: una zona di destinazione costituisce una base infrastrutturale e di sicurezza coerente, che consente all'organizzazione di sviluppare, lanciare e implementare rapidamente carichi di lavoro. Puoi utilizzare una zona di destinazione AWS Control Tower personalizzata per orchestrare il tuo ambiente.
3. Definizione di guardrail: implementa guardrail di sicurezza coerenti per il tuo ambiente mediante la tua zona di destinazione. AWS Control Tower fornisce un elenco di controlli obbligatori e facoltativi implementabili. I controlli obbligatori vengono implementati in automatico in caso di utilizzo di Control Tower. Esamina l'elenco dei controlli altamente consigliati e facoltativi e adotta quelli più adatti alle tue esigenze.
4. Limita l'accesso alle regioni appena aggiunte: per le nuove regioni Regioni AWS, IAM risorse come utenti e ruoli vengono propagate solo alle regioni specificate. Questa azione può essere eseguita tramite la console quando si utilizza Control Tower o modificando le politiche di IAM autorizzazione in AWS Organizations.
5. StackSets Considera AWS CloudFormation StackSets: aiutarti a distribuire risorse tra cui IAM politiche, ruoli e gruppi in diverse Account AWS regioni a partire da un modello approvato.

Risorse

Best practice correlate:

- SEC02-BP04 Affidati a un provider di identità centralizzato

Documenti correlati:

- AWS Control Tower
- Linee guida AWS sugli audit di sicurezza
- IAMLe migliori pratiche
- CloudFormation StackSets Utilizzato per fornire risorse in più aree Account AWS geografiche
- Organizzazioni FAQ
- AWS Organizations terminologia e concetti
- Migliori pratiche per le politiche di controllo dei servizi in un ambiente con AWS Organizations più account
- AWS Account Management Reference Guide
- Organizzazione AWS dell'ambiente utilizzando più account

Video correlati:

- Organizing Your AWS Environment Using Multiple Accounts
- Security Best Practices the Well-Architected Way
- Creazione e gestione di più account utilizzando AWS Control Tower
- Enable Control Tower for Existing Organizations

Workshop correlati:

- Control Tower Immersion Day

SEC01-BP02 Utente root e proprietà dell'account sicuro

L'utente root è l'utente con più privilegi in un account Account AWS, con accesso amministrativo completo a tutte le risorse all'interno dell'account e in alcuni casi non può essere vincolato dalle politiche di sicurezza. Disattivare l'accesso programmatico all'utente root, stabilire controlli appropriati

per l'utente root ed evitare l'uso di routine dell'utente root aiuta a ridurre il rischio di esposizione involontaria delle credenziali root e la conseguente compromissione dell'ambiente cloud.

Risultato desiderato: proteggere l'utente root riduce la possibilità di danni accidentali o intenzionali dovuti all'uso improprio delle credenziali dell'utente root. La creazione di controlli investigativi può anche permettere di avvisare il personale appropriato quando vengono eseguite azioni utilizzando l'utente root.

Anti-pattern comuni:

- Utilizzo dell'utente root per attività diverse da quelle che richiedono le proprie credenziali.
- Nessun test dei piani di emergenza su base regolare per verificare il funzionamento di infrastrutture critiche, processi e personale durante un'emergenza.
- Analisi limitata al tipico flusso di accesso all'account, trascurando di considerare o testare metodi alternativi di ripristino dell'account.
- La gestioneDNS, i server di posta elettronica e i provider telefonici non fanno parte del perimetro di sicurezza critico, in quanto vengono utilizzati nel flusso di ripristino dell'account.

Vantaggi dell'adozione di questa best practice: proteggere l'accesso all'utente root aumenta la sicurezza circa controlli e audit delle azioni nell'account

Livello di rischio associato se questa best practice non fosse adottata: elevato

Guida all'implementazione

AWS offre molti strumenti per proteggere il tuo account. Tuttavia, poiché alcune di queste misure non sono attivate per impostazione predefinita, è necessario intervenire direttamente per implementarle. Queste raccomandazioni costituiscono i passi fondamentali per mettere in sicurezza il proprio Account AWS. Durante l'implementazione di questi passaggi, è importante creare un processo di valutazione e monitoraggio continuo dei controlli di sicurezza.

La prima volta che si crea un account Account AWS, si inizia con un'identità che ha accesso completo a tutti i AWS servizi e le risorse dell'account. Questa identità è chiamata utente Account AWS root. Puoi accedere come utente root utilizzando l'indirizzo e-mail e la password usati per creare l'account. A causa dell'accesso elevato concesso all'utente AWS root, è necessario limitare l'uso dell'utente AWS root per eseguire attività che lo richiedono specificamente. Le credenziali di accesso dell'utente root devono essere attentamente protette e l'autenticazione a più fattori (MFA) deve essere sempre utilizzata per l'utente root. Account AWS

Oltre al normale flusso di autenticazione per accedere all'utente root utilizzando un nome utente, una password e un dispositivo di autenticazione a più fattori (MFA), esistono flussi di ripristino dell'account per accedere all'utente Account AWS root che ha accesso all'indirizzo e-mail e al numero di telefono associati all'account. Pertanto, è altrettanto importante proteggere l'account e-mail dell'utente root a cui vengono inviati l'e-mail di recupero e il numero di telefono associato all'account. Considerate anche le potenziali dipendenze circolari in cui l'indirizzo e-mail associato all'utente root è ospitato su server di posta elettronica o su risorse del servizio di nomi di dominio (DNS) dello stesso. Account AWS

Quando si utilizza AWS Organizations, ce ne sono più di uno, Account AWS ognuno dei quali ha un utente root. Un account è designato come account di gestione e sotto l'account di gestione è possibile aggiungere diversi livelli di account membri. La priorità è proteggere l'utente root dell'account di gestione, quindi occuparsi degli utenti root degli account membri. La strategia per la protezione dell'utente root dell'account di gestione può essere diversa da quella degli utenti root degli account membri ed è possibile effettuare controlli di sicurezza preventivi sugli utenti root degli account membri.

Passaggi dell'implementazione

Per stabilire i controlli per l'utente root, si consigliano i seguenti passaggi di implementazione. Ove applicabile, le raccomandazioni sono riferite alla versione 1.4.0 del benchmark CIS AWS Foundations. Oltre a questi passaggi, consulta le linee guida sulle AWS migliori pratiche per proteggere le tue risorse. Account AWS

Controlli preventivi

1. Imposta informazioni di contatto precise per l'account.
    a. Queste informazioni vengono utilizzate per il flusso di recupero della password smarrita, per il flusso di recupero dell'account MFA del dispositivo smarrito e per le comunicazioni critiche relative alla sicurezza con il team.
    b. Utilizza un indirizzo e-mail ospitato dal dominio aziendale, preferibilmente una lista di distribuzione, come indirizzo e-mail dell'utente root. L'utilizzo di una lista di distribuzione anziché l'account e-mail di un singolo individuo offre una maggiore ridondanza e continuità di accesso all'account root per lunghi periodi di tempo.
    c. Il numero di telefono indicato nelle informazioni di contatto deve essere dedicato e sicuro per questo scopo. Il numero di telefono non deve essere indicato o condiviso con nessuno.
2. Non creare chiavi di accesso per l'utente root. Se esistono chiavi di accesso, rimuovile (1.4)CIS.

a. Elimina le credenziali programmatiche a lunga durata (chiavi di accesso e segrete) per l'utente root.

b. Se esistono già chiavi di accesso utente root, è necessario passare i processi che utilizzano tali chiavi all'utilizzo di chiavi di accesso temporanee da un ruolo AWS Identity and Access Management (IAM), quindi eliminare le chiavi di accesso dell'utente root.

3. Stabilisci se è necessario memorizzare le credenziali per l'utente root.

   a. Se si utilizza AWS Organizations per creare nuovi account membro, la password iniziale per l'utente root sui nuovi account membro viene impostata su un valore casuale che non è esposto all'utente. Valuta la possibilità di utilizzare il flusso di reimpostazione della password del tuo account di gestione AWS dell'organizzazione per accedere all'account del membro, se necessario.

   b. Per l'account AWS aziendale autonomo Account AWS o di gestione, valuta la possibilità di creare e archiviare in modo sicuro le credenziali per l'utente root. Utilizzare MFA per l'utente root.

4. Utilizza i controlli preventivi per gli utenti root degli account membro in ambienti con AWS più account.

   a. Prendi in considerazione l'utilizzo del guardrail preventivo Disallow Creation of Root Access Keys for Root User per gli account membri.

   b. Prendi in considerazione l'utilizzo del guardrail preventivo Disallow Actions as a Root User per gli account membri.

5. Se sono necessarie le credenziali per l'utente root:

   a. Utilizza una password complessa.

   b. Attiva l'autenticazione a più fattori (MFA) per l'utente root, in particolare per gli account di AWS Organizations gestione (paganti) (1.5). CIS

   c. Prendi in considerazione MFA i dispositivi hardware per garantire resilienza e sicurezza, poiché i dispositivi monouso possono ridurre le possibilità che i dispositivi contenenti i tuoi MFA codici vengano riutilizzati per altri scopi. Verifica che MFA i dispositivi hardware alimentati da una batteria vengano sostituiti regolarmente. (CIS1.6)

      - Per eseguire la configurazione MFA per l'utente root, segui le istruzioni per la creazione di un MFAdispositivo virtuale MFA o hardware.

   d. Prendi in considerazione la possibilità di registrare più MFA dispositivi per il backup. Sono consentiti fino a 8 MFA dispositivi per account.

- Tieni presente che la registrazione di più di un MFA dispositivo per l'utente root disattiva automaticamente il processo di ripristino dell'account in caso di smarrimento del MFA dispositivo.

e. Conserva la password in modo sicuro e considera le dipendenze circolari se la password viene conservata elettronicamente. Non archiviate la password in modo tale da richiedere l'accesso alla stessa per Account AWS ottenerla.

6. Facoltativo: valuta la possibilità di stabilire un programma di rotazione periodica delle password per l'utente root.
   - Le best practice per la gestione delle credenziali dipendono dai requisiti normativi e di policy. Gli utenti root protetti da non MFA fanno affidamento sulla password come singolo fattore di autenticazione.
   - La modifica periodica della password dell'utente root riduce il rischio di utilizzo improprio di una password esposta inavvertitamente.

Controlli di rilevamento

- Crea allarmi per rilevare l'uso delle credenziali root (CIS1.7). Amazon GuardDuty può monitorare e inviare avvisi sull'utilizzo delle API credenziali degli utenti root tramite la RootCredentialUsagericerca.
- Valuta e implementa i controlli investigativi inclusi nel pacchetto di conformitàAWS Well-Architected Security Pillar AWS Config per, o se AWS Control Tower utilizzi, i controlli fortemente consigliati disponibili all'interno di Control Tower.

Guida operativa

- Stabilisci chi nell'organizzazione deve avere accesso alle credenziali dell'utente root.
  - Utilizza una regola per due persone in modo che nessun individuo abbia accesso a tutte le credenziali necessarie e ottenga l'accesso come utente root. MFA
  - Verifica che l'organizzazione, e non un singolo individuo, mantenga il controllo sul numero di telefono e sull'alias e-mail associati all'account (utilizzati per la reimpostazione e MFA il flusso di reimpostazione della password).
- Usa l'utente root solo per eccezione (CIS1.7).

  - L'utente AWS root non deve essere utilizzato per attività quotidiane, nemmeno amministrative. Effettua l'accesso come utente root solo per eseguire attivitàAWS che richiedono l'utente root. Tutte le altre azioni devono essere eseguite da altri utenti che assumono i ruoli appropriati.
- Verifica periodicamente che l'accesso all'utente root sia funzionante, in modo da testare le procedure prima di una situazione di emergenza che richieda l'uso delle credenziali dell'utente root.
- Verifica a intervalli regolari il funzionamento dell'indirizzo e-mail associato all'account e quelli indicati nei contatti alternativi. Monitora queste caselle di posta elettronica per le notifiche di sicurezza che potresti ricevere da <abuse@amazon.com>. Assicurati inoltre che i numeri di telefono associati all'account siano attivi.
- Prepara procedure di risposta agli incidenti per rispondere all'uso improprio dell'account root. Consulta la AWS Security Incident Response Guide e le best practice nella sezione Risposta agli imprevisti del whitepaper sul pilastro della sicurezza per ulteriori informazioni circa la creazione di una strategia di risposta agli incidenti adatta al tuo Account AWS.

Risorse

Best practice correlate:

- SEC01-BP01 Separare i carichi di lavoro utilizzando gli account
- SEC02-BP01 Usa meccanismi di accesso avanzati
- SEC03-BP02 Concedi l'accesso con privilegi minimi
- SEC03-BP03 Stabilire un processo di accesso di emergenza
- SEC10-BP05 Accesso preliminare alla fornitura

Documenti correlati:

- AWS Control Tower
- Linee guida AWS sugli audit di sicurezza
- IAMLe migliori pratiche
- Amazon GuardDuty : avviso di utilizzo delle credenziali root
- Una tep-by-step guida sul monitoraggio dell'uso delle credenziali root tramite CloudTrail
- MFAtoken approvati per l'uso con AWS
- Implementazione dell'accesso Break Glass su AWS
- I 10 migliori elementi di sicurezza da migliorare nel tuo Account AWS

- What do I do if I notice unauthorized activity in my Account AWS?

Video correlati:

- Organizing Your AWS Environment Using Multiple Accounts
- Security Best Practices the Well-Architected Way
- Limitazione dell'uso delle credenziali AWS root di AWS re:inforce 2022 — Migliori pratiche di sicurezza con AWS IAM

Esempi e lab correlati:

- Lab: configurazione e utente root Account AWS

SEC01-BP03 Identificare e convalidare gli obiettivi di controllo

In base ai requisiti di conformità e ai rischi identificati dal modello di rischio, individua e convalida gli obiettivi di controllo e i controlli da applicare al carico di lavoro. La convalida continua degli obiettivi di controllo e dei controlli aiuta a misurare l'efficacia della mitigazione dei rischi.

Risultato desiderato: gli obiettivi di controllo della sicurezza della tua azienda sono ben definiti e in linea con i requisiti di conformità. I controlli vengono implementati e applicati attraverso l'automazione e le policy e vengono costantemente valutati per verificarne l'efficacia nel raggiungimento degli obiettivi. Le prove dell'efficacia, sia in un determinato momento che in un determinato periodo di tempo, sono prontamente comunicate ai revisori.

Anti-pattern comuni:

- I requisiti normativi, le aspettative del mercato e gli standard di settore per una sicurezza certa non sono ben compresi dalla tua azienda.
- I framework di sicurezza informatica e gli obiettivi di controllo non sono allineati ai requisiti dell'azienda.
- L'implementazione dei controlli non è perfettamente allineata agli obiettivi di controllo in modo misurabile.
- L'automazione non viene utilizzata per creare report sull'efficacia dei tuoi controlli.

Livello di rischio associato se questa best practice non fosse adottata: elevato

Guida all'implementazione

I framework di sicurezza informatica comunemente utilizzati sono molti e possono costituire la base per gli obiettivi di controllo della sicurezza. Per determinare quale sia il framework più adatto alle tue esigenze, considera i requisiti normativi, le aspettative del mercato e gli standard di settore dell'azienda. Gli esempi includono AICPASOC2, PCI- HITRUST, ISO27001 e SP DSS 800-53. NIST

Per quanto riguarda gli obiettivi di controllo che identificate, cercate di capire in che modo AWS i servizi che utilizzate vi aiutano a raggiungerli. AWS ArtifactUtilizzateli per trovare documentazione e report in linea con i vostri framework di riferimento, che descrivano l'ambito di responsabilità coperto AWS e linee guida per il restante ambito di vostra responsabilità. Per ulteriori indicazioni specifiche sui servizi in linea con le varie dichiarazioni di controllo del framework, consulta le AWS Customer Compliance Guides.

Nel definire i controlli che raggiungono i tuoi obiettivi, codifica l'applicazione utilizzando i controlli preventivi e automatizza le mitigazioni mediante i controlli di rilevamento. Contribuite a prevenire configurazioni e azioni delle risorse non conformi durante l'utilizzo delle politiche di controllo del servizio (). AWS Organizations SCP Implementa le regole in AWS Config al fine di monitorare e segnalare le risorse non conformi, quindi passa a un modello di applicazione delle regole una volta che sei sicuro del loro comportamento. Per implementare set di regole predefinite e gestite in linea con i tuoi framework di sicurezza informatica, prendi in considerazione l'uso degli standard AWS Security Hub come prima opzione. Lo standard AWS Foundational Service Best Practices (FSBP) e il CIS AWS Foundations Benchmark sono buoni punti di partenza con controlli allineati a molti obiettivi condivisi tra più framework standard. Se Security Hub non dispone a livello intrinseco dei rilevamenti di controllo desiderati, è possibile integrarlo mediante i pacchetti di conformitàAWS Config.

Utilizzate i APNPartner Bundle consigliati dal team di AWS Global Security and Compliance Acceleration (GSCA) per ottenere assistenza da consulenti di sicurezza, agenzie di consulenza, sistemi di raccolta e rendicontazione delle prove, revisori e altri servizi complementari, quando necessario.

Passaggi dell'implementazione

1. Valuta i framework di sicurezza informatica comuni e allinea i tuoi obiettivi di controllo a quelli scelti.
2. Ottieni la documentazione pertinente sulle linee guida e le responsabilità per l'utilizzo del framework. AWS Artifact Comprendete quali aspetti della conformità rientrano nel modello di responsabilità condivisa e quali sono di vostra competenza. AWS

3. Utilizza politiche relative alle risorseSCPs, politiche di fiducia dei ruoli e altri ostacoli per prevenire configurazioni e azioni delle risorse non conformi.
4. Valuta l'implementazione degli standard e dei pacchetti di AWS Config conformità di Security Hub in linea con i tuoi obiettivi di controllo.

Risorse

Best practice correlate:

- SEC03-BP01 Definire i requisiti di accesso
- SEC04-BP01 Configurare la registrazione dei servizi e delle applicazioni
- SEC07-BP01 Comprendi il tuo schema di classificazione dei dati
- OPS01-BP03 Valuta i requisiti di governance
- OPS01-BP04 Valuta i requisiti di conformità
- PERF01-BP05 Utilizza politiche e architetture di riferimento
- COST02-BP01 Sviluppa politiche basate sui requisiti della tua organizzazione

Documenti correlati:

- AWS Customer Compliance Guides

Strumenti correlati:

- AWS Artifact

SEC01-BP04 Rimani aggiornato sulle minacce alla sicurezza e sui consigli

Rimani aggiornato sulle minacce più recenti e sulle misure di mitigazione monitorando le pubblicazioni di intelligence sulle minacce del settore e i feed di dati per gli aggiornamenti. Valuta le offerte di servizi gestiti che si aggiornano in automatico in base ai dati sulle minacce più recenti.

Risultato desiderato: rimani informato mentre le pubblicazioni di settore si aggiornano con le ultime minacce e raccomandazioni. L'automazione viene utilizzata per rilevare potenziali vulnerabilità ed esposizioni man mano che si identificano nuove minacce. Intraprendi azioni di mitigazione contro queste minacce. Adottate AWS servizi che si aggiornano automaticamente con le più recenti informazioni sulle minacce.

Anti-pattern comuni:

- Non disporre di un meccanismo affidabile e ripetibile per rimanere informati sulle ultime informazioni sulle minacce.
- Mantenere un inventario manuale del portafoglio tecnologico, dei carichi di lavoro e delle dipendenze che richiedono un esame umano per individuare potenziali vulnerabilità ed esposizioni.
- Non disporre di meccanismi per aggiornare i carichi di lavoro e le dipendenze alle ultime versioni disponibili, che forniscono mitigazioni note delle minacce.

Vantaggi dell'adozione di questa best practice: l'utilizzo di fonti di intelligence sulle minacce per rimanere aggiornati riduce il rischio di lasciarsi sfuggire importanti cambiamenti nel panorama delle minacce in grado di pregiudicare la tua azienda. L'automazione in atto per scansionare, rilevare e correggere eventuali vulnerabilità o esposizioni nei carichi di lavoro e nelle relative dipendenze può aiutarti a mitigare i rischi in modo rapido e prevedibile, rispetto alle alternative manuali. In questo modo puoi controllare i tempi e i costi relativi alla mitigazione delle vulnerabilità.

Livello di rischio associato se questa best practice non fosse adottata: elevato

Guida all'implementazione

Consulta le pubblicazioni di intelligence sulle minacce per costanti aggiornamenti sul panorama delle minacce. Consultate la knowledge base MITREATT&CK per la documentazione sulle tattiche, le tecniche e le procedure antagonistiche note (). TTPs Consulta MITRE l'elenco delle vulnerabilità e delle esposizioni comuni (CVE) per rimanere informato sulle vulnerabilità note nei prodotti su cui fai affidamento. Comprendi i rischi critici per le applicazioni web con il popolare OWASP progetto Top 10 dell'Open Worldwide Application Security Project (OWASP).

Rimani aggiornato sugli eventi di AWS sicurezza e sulle procedure di correzione consigliate con AWS Security Bulletins for. CVEs

Per ridurre l'impegno complessivo e il sovraccarico legati all'aggiornamento, prendi in considerazione l'utilizzo di AWS servizi che incorporano automaticamente nuove informazioni sulle minacce nel tempo. Ad esempio, Amazon si GuardDuty aggiorna con le informazioni sulle minacce del settore per rilevare comportamenti anomali e firme di minacce all'interno dei tuoi account. Amazon Inspector mantiene automaticamente aggiornato un database dei CVEs dati utilizzati per le sue funzionalità di scansione continua. AWS WAF e AWS Shield Advanced forniscono gruppi di regole gestiti, aggiornati in automatico all'emergere di nuove minacce.

Esamina il pilastro dell'eccellenza operativa Well-Architected per la gestione e l'applicazione di patch automatizzate del parco.

Passaggi dell'implementazione

- Abbonati agli aggiornamenti per le pubblicazioni di intelligence sulle minacce pertinenti alla tua azienda e al tuo settore. Abbonati ai bollettini sulla sicurezza AWS .
- Prendi in considerazione l'adozione di servizi che incorporano automaticamente nuove informazioni sulle minacce, come Amazon GuardDuty e Amazon Inspector.
- Implementa una strategia di gestione e applicazione delle patch del parco in linea con le best practice del pilastro dell'eccellenza operativa Well-Architected.

Risorse

Best practice correlate:

- SEC01-BP07 Identifica le minacce e dai priorità alle mitigazioni utilizzando un modello di minaccia
- OPS01-BP05 Valuta il panorama delle minacce
- OPS11-BP01 Adottate un processo per il miglioramento continuo

SEC01-BP05 Ridurre l'ambito di gestione della sicurezza

Determina se puoi ridurre l'ambito di sicurezza utilizzando AWS servizi che spostano la gestione di determinati controlli su AWS (servizi gestiti). Questi servizi possono contribuire a ridurre le attività di manutenzione della sicurezza, come il provisioning dell'infrastruttura, l'impostazione del software, il patching o i backup.

Risultato desiderato: quando si selezionano i AWS servizi per il carico di lavoro, si considera l'ambito della gestione della sicurezza. Il costo delle spese generali di gestione e delle attività di manutenzione (il costo totale di proprietà oTCO) viene confrontato con il costo dei servizi selezionati, oltre ad altre considerazioni di Well-Architected. La documentazione relativa al AWS controllo e alla conformità viene incorporata nelle procedure di valutazione e verifica del controllo.

Anti-pattern comuni:

- Implementazione dei carichi di lavoro senza comprendere a fondo il modello di responsabilità condivisa per i servizi selezionati.

- Hosting di database e altre tecnologie su macchine virtuali senza aver valutato un servizio gestito equivalente.
- Mancata inclusione delle attività di gestione della sicurezza nel costo totale di proprietà delle tecnologie di hosting su macchine virtuali rispetto alle opzioni di servizio gestito.

Vantaggi dell'adozione di questa best practice: l'utilizzo di servizi gestiti può ridurre l'onere complessivo della gestione dei controlli operativi della sicurezza, così da ridurre rischi per la sicurezza e costo totale di proprietà. Il tempo che altrimenti sarebbe dedicato a determinate attività di sicurezza può essere reinvestito in attività che forniscono maggior valore alla tua azienda. I servizi gestiti possono anche ridurre l'ambito dei requisiti di conformità spostando alcuni requisiti di controllo su AWS.

Livello di rischio associato se questa best practice non fosse adottata: medio

Guida all'implementazione

Le modalità di integrazione dei componenti del carico di lavoro su AWS sono molteplici. L'installazione e l'esecuzione di tecnologie su EC2 istanze Amazon spesso richiedono l'assunzione della maggior parte della responsabilità complessiva in materia di sicurezza. Per contribuire a ridurre l'onere derivante dall'utilizzo di determinati controlli, AWS individua i servizi gestiti che riducano la portata del modello di responsabilità condivisa e comprendi come utilizzarli nell'architettura esistente. Gli esempi includono l'utilizzo di Amazon Relational Database Service (RDSAmazon) per la distribuzione di database, Amazon Elastic Kubernetes Service (Amazon) o Amazon Elastic ECS Container EKS Service(Amazon) per orchestrare contenitori o l'utilizzo di opzioni serverless. Quando sviluppi nuove applicazioni, pensa a quali servizi possono contribuire a ridurre i tempi e i costi di implementazione e gestione dei controlli di sicurezza.

Anche i requisiti di conformità possono essere un fattore di scelta dei servizi. I servizi gestiti possono spostare la conformità di alcuni requisiti a. AWS Parlate con il vostro team addetto alla conformità in merito alla loro capacità di controllare gli aspetti dei servizi che gestite e gestite e di accettare le dichiarazioni di controllo nei rapporti di AWS audit pertinenti. Potete fornire agli auditor o AWS Artifactalle autorità di regolamentazione gli elementi degli audit presenti come prova dei controlli di sicurezza. AWS Puoi anche utilizzare le linee guida sulla responsabilità fornite da alcuni degli elementi di AWS audit per progettare la tua architettura, insieme alle Customer Compliance Guides.AWS Queste indicazioni aiutano a determinare i controlli di sicurezza aggiuntivi da mettere in atto per supportare i casi d'uso specifici del sistema.

Quando utilizzi servizi gestiti, acquisisci familiarità con il processo di aggiornamento delle risorse alle versioni più recenti (ad esempio, l'aggiornamento della versione di un database gestito da Amazon RDS o il runtime di un linguaggio di programmazione per una AWS Lambda funzione). Anche se il servizio gestito può eseguire questa operazione per tuo conto, la configurazione della tempistica dell'aggiornamento e la conoscenza dell'impatto sulle tue operazioni restano di tua responsabilità. Strumenti come AWS Health ti consentono di tracciare e gestire questi aggiornamenti in tutti i tuoi ambienti.

Passaggi dell'implementazione

1. Valuta i componenti del tuo carico di lavoro sostituibili con un servizio gestito.
    a. Se stai migrando un carico di lavoro verso AWS, prendi in considerazione la riduzione della gestione (tempo e spese) e la riduzione del rischio quando valuti se riospitare, rifattorizzare, ripiattaforma, ricostruire o sostituire il carico di lavoro. A volte un investimento aggiuntivo all'inizio di una migrazione può comportare risparmi significativi nel lungo periodo.
2. Prendi in considerazione l'implementazione di servizi gestitiRDS, come Amazon, anziché installare e gestire le tue implementazioni tecnologiche.
3. Utilizza le linee guida sulla responsabilità riportate AWS Artifact di seguito per determinare i controlli di sicurezza da adottare per il tuo carico di lavoro.
4. Tenete un inventario delle risorse in uso e continuate a up-to-date utilizzare nuovi servizi e approcci per identificare nuove opportunità per ridurre l'ambito di applicazione.

Risorse

Best practice correlate:

- PERF02-BP01 Seleziona le migliori opzioni di elaborazione per il tuo carico di lavoro
- PERF03-BP01 Utilizza un data store appositamente progettato che supporti al meglio i requisiti di accesso e archiviazione dei dati
- SUS05-BP03 Utilizza servizi gestiti

Documenti correlati:

- Eventi del ciclo di vita pianificati per AWS Health

Strumenti correlati:

- AWS Health
- AWS Artifact
- AWS Customer Compliance Guides

Video correlati:

- Come posso migrare a un'istanza Amazon RDS o Aurora SQL My DB utilizzando? AWS DMS
- AWS re:Invent 2023 - Gestisci gli eventi del ciclo di vita delle risorse su larga scala con AWS Health

SEC01-BP06 Automatizza l'implementazione dei controlli di sicurezza standard

Applica DevOps pratiche moderne mentre sviluppi e distribuisci controlli di sicurezza standard in tutti i tuoi ambienti. AWS Definisci controlli e configurazioni di sicurezza standard utilizzando modelli Infrastructure as Code (IaC), acquisisci le modifiche in un sistema di controllo delle versioni, testa le modifiche come parte di una pipeline CI/CD e automatizza l'implementazione delle modifiche ai tuoi ambienti. AWS

Risultato desiderato: i modelli IaC acquisiscono controlli di sicurezza standardizzati, inserendoli in un sistema di controllo delle versioni. Le pipeline CI/CD consentono di rilevare le modifiche e automatizzare i test e l'implementazione degli ambienti. AWS Sono presenti guardrail per rilevare e fornire avvisi in caso di configurazioni errate nei modelli prima di procedere all'implementazione. I carichi di lavoro vengono implementati in ambienti dotati di controlli standard. I team hanno accesso all'implementazione di configurazioni di servizio approvate tramite un meccanismo self-service. Sono disponibili strategie di backup e ripristino sicure per le configurazioni di controllo, gli script e i dati correlati.

Anti-pattern comuni:

- Apportare modifiche ai controlli di sicurezza standard manualmente, tramite una console Web o un'interfaccia a riga di comando.
- Affidarsi ai singoli team del carico di lavoro per implementare manualmente i controlli definiti da un team centrale.
- Affidarsi a un team di sicurezza centrale per implementare i controlli a livello di carico di lavoro su richiesta di un team del carico di lavoro.

- Consentire agli stessi individui o team di sviluppare, testare e implementare script di automazione per il controllo della sicurezza senza un'adeguata separazione dei compiti o dei controlli e degli equilibri.

Vantaggi dell'adozione di questa best practice: l'utilizzo di modelli per definire i controlli di sicurezza standard consente di tracciare e confrontare le modifiche nel tempo con un sistema di controllo delle versioni. L'uso dell'automazione per testare e implementare le modifiche crea standardizzazione e prevedibilità, aumentando le possibilità di una corretta implementazione e riducendo le attività manuali ripetitive. Fornire un meccanismo self-service per consentire ai team addetti al carico di lavoro di implementare servizi e configurazioni approvati riduce il rischio di configurazioni errate e usi impropri. Questo li aiuta anche a incorporare i controlli nelle prime fasi del processo di sviluppo.

Livello di rischio associato se questa best practice non fosse adottata: medio

Guida all'implementazione

Se si seguono le pratiche descritte in SEC01-BP01 Separazione dei carichi di lavoro utilizzando gli account, si ottengono più unità per ambienti diversi da gestire. Account AWS AWS Organizations Sebbene ciascuno di questi ambienti e carichi di lavoro possa richiedere controlli di sicurezza distinti, puoi standardizzarne alcuni in tutta l'organizzazione. Gli esempi includono l'integrazione di gestori dell'identità digitale centralizzati, la definizione di reti e firewall e la configurazione di posizioni standard per l'archiviazione e l'analisi dei log. Allo stesso modo in cui puoi utilizzare infrastructure as code (IaC) per applicare lo stesso criterio dello sviluppo del codice dell'applicazione al provisioning dell'infrastruttura, puoi usare l'IaC anche per definire e implementare controlli di sicurezza standard.

Se possibile, definisci i controlli di sicurezza in modo dichiarativo, ad esempio in AWS CloudFormation, e archiviali in un sistema di controllo del codice sorgente. Utilizza DevOps procedure per automatizzare l'implementazione dei controlli per versioni più prevedibili, esegui test automatici utilizzando strumenti come AWS CloudFormation Guarde rilevando eventuali differenze tra i controlli implementati e la configurazione desiderata. Puoi utilizzare servizi come AWS CodePipeline, AWS CodeBuild e AWS CodeDeploy per creare una pipeline CI/CD. Prendi in considerazione le indicazioni contenute nella sezione Organizzazione AWS dell'ambiente utilizzando più account per configurare questi servizi nei rispettivi account, separati dalle altre pipeline di distribuzione.

È inoltre possibile definire modelli per standardizzare la definizione e la distribuzione Account AWS, i servizi e le configurazioni. Questa tecnica consente a un team di sicurezza centrale di gestire queste

definizioni e di fornirle ai team che si occupano dei carichi di lavoro attraverso un approccio self-service. Un modo per raggiungere questo obiettivo è utilizzare Service Catalog, dove è possibile pubblicare modelli come prodotti che i team addetti al carico di lavoro possono integrare nelle proprie implementazioni della pipeline. AWS Control Tower offre alcuni modelli e controlli come punto di partenza. Control Tower offre anche la funzionalità Account Factory, che consente ai team addetti al carico di lavoro di creare di nuovi Account AWS mediante gli standard definiti da te. Questa funzionalità aiuta a rimuovere le dipendenze da un team centrale per l'approvazione e la creazione di nuovi account quando vengono identificati come necessari dai team del carico di lavoro. Potresti aver bisogno di questi account per isolare i diversi componenti del carico di lavoro in base a motivi quali la funzione che svolgono, la sensibilità dei dati elaborati o il loro comportamento.

Passaggi dell'implementazione

1. Determina come archivierai e manterrai i tuoi modelli in un sistema di controllo delle versioni.
2. Crea pipeline CI/CD per testare e implementare i tuoi modelli. Definisci i test per verificare che non ci siano configurazioni errate e che i modelli siano conformi agli standard aziendali.
3. Crea un catalogo di modelli standardizzati da distribuire ai team addetti ai carichi di lavoro Account AWS e di servizi in base alle tue esigenze.
4. Implementa strategie di backup e ripristino sicure per le configurazioni di controllo, gli script e i dati correlati.

Risorse

Best practice correlate:

- OPS05-BP01 Usa il controllo della versione
- OPS05-BP04 Utilizza sistemi di gestione della compilazione e dell'implementazione
- REL08-BP05 Implementa le modifiche con l'automazione
- SUS06-BP01 Adotta metodi in grado di introdurre rapidamente miglioramenti alla sostenibilità

Documenti correlati:

- Organizzazione dell'ambiente utilizzando più account AWS

Esempi correlati:

- Automatizza la creazione di account e il provisioning delle risorse utilizzando Service Catalog e AWS OrganizationsAWS Lambda
- Rafforza la DevOps pipeline e proteggi i dati con AWS Secrets Manager, e AWS KMSAWS Certificate Manager

Strumenti correlati:

- AWS CloudFormation Guard
- Landing Zone Accelerator attivo AWS

SEC01-BP07 Identifica le minacce e dai priorità alle mitigazioni utilizzando un modello di minaccia

Esegui la modellazione delle minacce per identificare e mantenere un up-to-date registro delle potenziali minacce e delle relative mitigazioni per il tuo carico di lavoro. Definisci le priorità delle minacce e adatta le mitigazioni dei controlli di sicurezza per prevenire, intercettare e rispondere. Riesamina e mantieni questo aspetto nel contesto del tuo carico di lavoro e dell'evoluzione del panorama della sicurezza.

Livello di rischio associato se questa best practice non fosse adottata: elevato

Guida all'implementazione

Che cos'è la modellazione delle minacce?

"La modellazione delle minacce mira a identificare, comunicare e comprendere minacce e mitigazioni nel contesto della protezione di qualcosa di valore". — Modellazione delle minacce delle applicazioni Open Web Application Security Project (OWASP)

Perché adottare la modellazione delle minacce?

I sistemi sono complessi, e nel tempo lo diventano sempre di più, e capaci di fornire un maggiore valore aziendale e una maggiore soddisfazione e coinvolgimento dei clienti. Ciò significa che le decisioni di progettazione IT devono tenere conto di un numero sempre maggiore di casi d'uso. Questa complessità e il numero di combinazioni di casi d'uso rendono in genere gli approcci non strutturati inefficaci per individuare e mitigare le minacce. È invece necessario un approccio sistematico per enumerare le potenziali minacce al sistema ed elaborare le mitigazioni, oltre che per stabilirne le priorità per assicurarsi che le risorse limitate dell'organizzazione abbiano il massimo impatto nel migliorare lo stato di sicurezza complessiva del sistema.

La modellazione delle minacce è progettata per offrire questo approccio sistematico, con l'obiettivo di trovare e affrontare i problemi nelle prime fasi del processo di progettazione, quando le mitigazioni hanno un costo e un impegno relativi bassi rispetto alle fasi successive del ciclo di vita. Questo approccio è in linea con il principio di sicurezza shift-left. In definitiva, la modellazione delle minacce si integra con il processo di gestione del rischio di un'organizzazione e aiuta a prendere decisioni sui controlli da implementare utilizzando un approccio orientato alle minacce.

Quando eseguire la modellazione delle minacce?

La modellazione delle minacce deve essere avviata il prima possibile nel ciclo di vita del carico di lavoro, in modo da avere una maggiore flessibilità di intervento sulle minacce identificate. Come per i bug del software, prima si identificano le minacce, più è conveniente affrontarle. Un modello di minacce è un documento vivo e deve continuare a evolvere in base ai cambiamenti dei carichi di lavoro. I modelli di minaccia vanno riesaminati nel tempo, anche in caso di modifiche importanti, di cambiamenti nel panorama delle minacce o di adozione di nuove funzionalità o servizi.

Passaggi dell'implementazione

In che modo è possibile eseguire la modellazione delle minacce?

Esistono diversi modi per eseguire la modellazione delle minacce. Come per i linguaggi di programmazione, anche in questo caso ci sono vantaggi e svantaggi e bisogna scegliere il metodo più adatto alle proprie esigenze. Un approccio consiste nell'iniziare con 4 Question Frame for Threat Modeling di Shostack, che pone domande aperte per fornire una struttura per il tuo esercizio di modellazione delle minacce:

1. A cosa si sta lavorando?

   Questa domanda ha lo scopo di aiutare a comprendere e concordare il sistema che si sta costruendo e i dettagli di tale sistema che sono rilevanti per la sicurezza. La creazione di un modello o di un diagramma è la soluzione più comune per rispondere a questa domanda, in quanto consente di visualizzare ciò che si sta creando, ad esempio utilizzando un diagramma di flusso dei dati. Scrivere ipotesi e dettagli importanti del sistema aiuta anche a definire l'ambito di applicazione. Ciò consente a tutti coloro che contribuiscono al modello di minaccia di concentrarsi sulla stessa cosa ed evitare lunghe deviazioni su out-of-scope argomenti (comprese le versioni obsolete del sistema). Ad esempio, se si sta realizzando un'applicazione Web, probabilmente non vale la pena procedere alla modellazione per la sequenza di avvio attendibile del sistema operativo per i browser client, poiché non si ha la possibilità di influire su questo aspetto con il proprio progetto.

2. Che cosa può andare storto?

   In questa fase si identificano le minacce al sistema. Le minacce sono azioni o eventi accidentali o intenzionali che producono impatti indesiderati e potrebbero compromettere la sicurezza del sistema. Senza una visione chiara di ciò che potrebbe andare storto, non è possibile fare nulla per evitarlo.

   Non esiste un elenco canonico di ciò che può andare storto. La creazione di questo elenco richiede un brainstorming e la collaborazione tra tutte le persone del team e le persone pertinenti coinvolte nell'esercizio di modellazione delle minacce. È possibile agevolare le proprie riflessioni utilizzando un modello per identificare le minacce, ad esempio suggerendo diverse categorie da valutare: contraffazione STRIDE, manomissione, ripudio, divulgazione di informazioni, negazione del servizio ed elevazione dei privilegi. Inoltre, potresti contribuire al brainstorming esaminando gli elenchi esistenti e facendo ricerche per trarne ispirazione, tra cui la Top 10, il Threat Catalog e il catalogo delle minacce della tua organizzazione. OWASP HiTrust

3. Cosa si intende fare al riguardo?

   Come nel caso della domanda precedente, non esiste un elenco canonico di tutte le possibili mitigazioni. Gli input di questa fase sono le minacce, gli attori e le aree di miglioramento identificate nella fase precedente.

   Sicurezza e conformità sono una responsabilità condivisa tra AWS e l'utente. È importante capire che quando si chiede "Che si farà al riguardo?", si chiede anche "Chi è responsabile? Chi ha la responsabilità di fare qualcosa?" Comprendere l'equilibrio delle responsabilità tra voi e AWS aiutarvi ad adattare l'esercizio di modellazione delle minacce alle mitigazioni che sono sotto il vostro controllo, che in genere sono una combinazione di opzioni di configurazione del AWS servizio e mitigazioni specifiche del sistema.

   Per quanto riguarda la AWS parte della responsabilità condivisa, scoprirete che i AWS servizi rientrano nell'ambito di molti programmi di conformità. Questi programmi ti aiutano a comprendere i solidi controlli in atto AWS per mantenere la sicurezza e la conformità del cloud. I rapporti di controllo di questi programmi possono essere scaricati per AWS i clienti da AWS Artifact.

   Indipendentemente dai AWS servizi utilizzati, esiste sempre un elemento di responsabilità del cliente e le mitigazioni in linea con queste responsabilità dovrebbero essere incluse nel modello di minaccia. Per mitigare il controllo di sicurezza AWS dei servizi stessi, è consigliabile prendere in considerazione l'implementazione di controlli di sicurezza su più domini, inclusi domini come la gestione delle identità e degli accessi (autenticazione e autorizzazione), la protezione dei

dati (a riposo e in transito), la sicurezza dell'infrastruttura, la registrazione e il monitoraggio. La documentazione di ogni AWS servizio contiene un capitolo dedicato alla sicurezza che fornisce indicazioni sui controlli di sicurezza da considerare come mitigazioni. È importante considerare il codice che si sta scrivendo e le sue dipendenze e pensare ai controlli attuabili per affrontare queste minacce. Questi controlli potrebbero corrispondere a elementi quali la convalida degli input, la gestione delle sessioni e la gestione dei limiti. Spesso la maggior parte delle vulnerabilità viene introdotta nel codice personalizzato, quindi è bene concentrarsi su quest'area.

4. È stato fatto un buon lavoro?

   L'obiettivo è il miglioramento da parte del team e dell'organizzazione sia della qualità dei modelli di minacce sia della relativa velocità di esecuzione nel tempo. Questi miglioramenti derivano da una combinazione di pratica, apprendimento, insegnamento e revisione. Per approfondire e sperimentare nella pratica, è consigliabile che tu e il tuo team completiate il corso di formazione Threat modeling the right way for builders training course o il workshop. Inoltre, se stai cercando indicazioni su come integrare la modellazione delle minacce nel ciclo di vita di sviluppo delle applicazioni della tua organizzazione, consulta il post How to approach threat modeling sul Security Blog. AWS

Threat Composer

Per aiutarti e guidarti nell'esecuzione della modellazione delle minacce, prendi in considerazione l'utilizzo dello strumento Threat Composer, che mira a ridurre i tempi di modellazione delle minacce. time-to-value Lo strumento consente di eseguire le seguenti operazioni:

- Scrivere dichiarazioni utili sulle minacce in linea con la sintassi delle minacce che funzionino in un flusso di lavoro naturale non lineare
- Generare un modello di minaccia leggibile dall'uomo
- Generare un modello di minaccia leggibile dal computer per consentire la gestione dei modelli di minaccia come codice
- Velocizzare l'individuazione delle aree di miglioramento della qualità e della copertura utilizzando l'area del pannello di controllo contenente le informazioni dettagliate

Per ulteriori informazioni, visita Threat Composer e passa all'area di lavoro esemplificativa definita dal sistema.

Risorse

Best practice correlate:

- SEC01-BP03 Identificare e convalidare gli obiettivi di controllo
- SEC01-BP04 Rimani aggiornato sulle minacce alla sicurezza e sui consigli
- SEC01-BP05 Ridurre l'ambito di gestione della sicurezza
- SEC01-BP08 Valuta e implementa regolarmente nuovi servizi e funzionalità di sicurezza

Documenti correlati:

- Come affrontare la modellazione delle minacce (Security Blog)AWS
- NIST: Guida alla modellazione delle minacce di sistema incentrata sui dati

Video correlati:

- AWS Summit ANZ 2021 - Come affrontare la modellizzazione delle minacce
- AWS Summit ANZ 2022 - Sicurezza scalabile: ottimizzazione per una consegna rapida e sicura

Formazione correlata:

- Modellazione delle minacce nel modo giusto per i costruttori: formazione autonoma virtuale di AWS Skill Builder
- Modellazione delle minacce nel modo giusto per i costruttori — Workshop AWS

Strumenti correlati:

- Threat Composer

SEC01-BP08 Valuta e implementa regolarmente nuovi servizi e funzionalità di sicurezza

Valuta e implementa servizi e funzionalità di sicurezza di AWS e AWS partner che ti aiutano a far evolvere il livello di sicurezza del tuo carico di lavoro.

Risultato desiderato: hai adottato una procedura standard che ti informa sulle nuove funzionalità e servizi rilasciati da e Partner. AWS AWS Puoi valutare come queste nuove funzionalità influenzino la progettazione di controlli attuali e nuovi per i tuoi ambienti e carichi di lavoro.

Anti-pattern comuni:

- Non ti iscrivi a AWS blog e RSS feed per conoscere rapidamente nuove funzionalità e servizi pertinenti
- Fai affidamento su notizie e aggiornamenti sui servizi e sulle funzioni di sicurezza provenienti da fonti di seconda mano
- Non incoraggiate AWS gli utenti della vostra organizzazione a tenersi informati sugli ultimi aggiornamenti

Vantaggi dell'adozione di questa best practice: rimanere aggiornati sui nuovi servizi e funzionalità di sicurezza, consente di adottare decisioni informate sull'implementazione dei controlli negli ambienti cloud e nei carichi di lavoro. Queste fonti aiutano a sensibilizzare l'opinione pubblica sull'evoluzione del panorama della sicurezza e su come AWS i servizi possono essere utilizzati per proteggersi da minacce nuove ed emergenti.

Livello di rischio associato se questa best practice non fosse adottata: basso

Guida all'implementazione

AWS informa i clienti sui nuovi servizi e funzionalità di sicurezza attraverso diversi canali:

- AWS Cosa c'è di nuovo
- AWS Blog di notizie
- AWS Security Blog
- AWS Security Bulletins
- AWS documentation overview

Puoi iscriverti a un argomento sugli aggiornamenti AWS giornalieri delle funzionalità utilizzando Amazon Simple Notification Service (AmazonSNS) per un riepilogo giornaliero completo degli aggiornamenti. Alcuni servizi di sicurezza, come Amazon GuardDuty e AWS Security Hub, forniscono i propri SNS argomenti per rimanere informati su nuovi standard, scoperte e altri aggiornamenti per quei servizi specifici.

Anche durante conferenze, eventi e webinar che si tengono ogni anno in tutto il mondo, vengono annunciati nuovi servizi e funzionalità. Segnaliamo in particolare conferenza annuale sulla sicurezza AWS re:Inforce e la conferenza più generale AWS re:Invent. I canali di AWS notizie menzionati in precedenza condividono questi annunci di conferenze sulla sicurezza e su altri servizi, e puoi

guardare le sessioni di approfondimento didattiche online sul canale Eventi all'indirizzo.AWS YouTube

Puoi anche chiedere al team del tuo Account AWS informazioni sugli aggiornamenti e consigli più recenti sui servizi di sicurezza. Puoi contattare il team tramite il modulo Sales Support se non disponi dei loro recapiti diretti. Allo stesso modo, se ti sei abbonato a AWS Enterprise Support, riceverai aggiornamenti settimanali dal tuo Technical Account Manager (TAM) e potrai programmare una riunione di revisione periodica con loro.

Passaggi dell'implementazione

1. Abbonatevi ai vari blog e bollettini con il vostro RSS lettore preferito o all'argomento Daily Features Updates. SNS
2. Valuta a quali AWS eventi partecipare per conoscere in prima persona nuove funzionalità e servizi.
3. Organizza riunioni con il tuo Account AWS team per qualsiasi domanda sull'aggiornamento dei servizi e delle funzionalità di sicurezza.
4. Prendi in considerazione la possibilità di abbonarti a Enterprise Support per avere consultazioni regolari con un Technical Account Manager ()TAM.

Risorse

Best practice correlate:

- PERF01-BP01 Scopri e comprendi i servizi e le funzionalità cloud disponibili
- COST01-BP07 Resta aggiornato sulle nuove release del servizio up-to-date

Gestione dell'identità e degli accessi

Questions

- SEC2. Come si gestisce l'autenticazione per persone e macchine?
- SEC3. Come si gestiscono le autorizzazioni per persone e macchine?

SEC2. Come si gestisce l'autenticazione per persone e macchine?

Esistono due tipi di identità che è necessario gestire quando si tratta di gestire carichi di lavoro sicuri. AWS Comprendere il tipo di identità necessaria per gestire e concedere l'accesso ti aiuta a verificare che le identità corrette abbiano accesso alle risorse giuste nelle condizioni adeguate.

Identità umane: gli amministratori, gli sviluppatori, gli operatori e gli utenti finali richiedono un'identità per accedere agli ambienti e alle applicazioni. AWS Si tratta di membri dell'organizzazione o utenti esterni con cui collabora e che interagiscono con AWS le risorse dell'utente tramite un browser Web, un'applicazione client o strumenti interattivi da riga di comando.

Identità delle macchine: le applicazioni di servizio, gli strumenti operativi e i carichi di lavoro richiedono un'identità per effettuare richieste ai AWS servizi, ad esempio per leggere i dati. Queste identità includono macchine in esecuzione nel tuo AWS ambiente, come EC2 istanze o AWS Lambda funzioni Amazon. Puoi gestire le identità di macchine anche per soggetti esterni che necessitano dell'accesso. Inoltre, potresti avere anche macchine esterne AWS che richiedono l'accesso al tuo AWS ambiente.

Best practice

- SEC02-BP01 Usa meccanismi di accesso avanzati
- SEC02-BP02 Usa credenziali temporanee
- SEC02-BP03 Archivia e utilizza i segreti in modo sicuro
- SEC02-BP04 Affidati a un provider di identità centralizzato
- SEC02-BP05 Verifica e ruota periodicamente le credenziali
- SEC02-BP06 Utilizza gruppi e attributi di utenti

SEC02-BP01 Usa meccanismi di accesso avanzati

Gli accessi (autenticazione tramite credenziali di accesso) possono presentare dei rischi quando non si utilizzano meccanismi come l'autenticazione a più fattori (MFA), specialmente in situazioni in cui le credenziali di accesso sono state divulgate inavvertitamente o sono facilmente intuibili. Utilizzate meccanismi di accesso efficaci per ridurre questi rischi richiedendo politiche rigorose in materia di password. MFA

Risultato desiderato: riduci i rischi di accesso involontario alle credenziali AWS utilizzando meccanismi di accesso efficaci per AWS Identity and Access Management (IAM) gli utenti, l'utente Account AWS root AWS IAM Identity Center(successore del AWS Single Sign-On) e i provider di identità di terze parti. Ciò significa richiedere MFA e applicare politiche complesse in materia di password e rilevare comportamenti di accesso anomali.

Anti-pattern comuni:

- Non applicare una politica solida in materia di password per le identità, comprese password complesse e. MFA
- Condivisione delle stesse credenziali tra utenti diversi.
- Nessun utilizzo di controlli investigativi per gli accessi sospetti.

Livello di rischio associato se questa best practice non fosse adottata: elevato

Guida all'implementazione

Ci sono molti modi in cui le identità umane possono accedere ad AWS. È consigliabile affidarsi a un AWS provider di identità centralizzato che utilizzi la federazione (federazione diretta o utilizzo AWS IAM Identity Center) per l'autenticazione. AWS In questo caso, è necessario stabilire un processo di accesso sicuro con il gestore dell'identità digitale o con Microsoft Active Directory.

La prima volta che si apre un Account AWS, si inizia con un Account AWS utente root. L'utente root dell'account va utilizzato solo per configurare l'accesso degli utenti (e per le attività che richiedono l'utente root). È importante attivare MFA l'account come utente root subito dopo aver aperto il proprio Account AWS e proteggere l'utente root utilizzando la guida alle AWS migliori pratiche.

Se crei utenti in AWS IAM Identity Center, proteggi la procedura di accesso in quel servizio. Per le identità dei consumatori, puoi utilizzare pool di utenti di Amazon Cognito e proteggere il processo di accesso del servizio in questione o utilizzare uno dei gestori dell'identità digitale supportati dai pool di utenti di Amazon Cognito.

Se utilizzi utenti AWS Identity and Access Management (IAM), proteggi il processo di accesso utilizzando. IAM

Indipendentemente dal metodo di accesso, è fondamentale applicare una policy di accesso efficace.

Passaggi dell'implementazione

Di seguito sono indicate raccomandazioni generali per l'accesso sicuro. Le impostazioni effettive che configuri devono essere impostate dalla politica aziendale o utilizzare uno standard come NIST800-63.

- Richiedere. MFA È una IAMbest practice da richiedere MFA per le identità umane e i carichi di lavoro. L'attivazione MFA offre un ulteriore livello di sicurezza che richiede agli utenti di fornire credenziali di accesso e una password monouso (OTP) o una stringa verificata e generata crittograficamente da un dispositivo hardware.
- Applica una lunghezza minima della password, fattore primario nell'efficacia della password.

- Applica la complessità delle password in modo che sia più difficile individuarle.
- Consenti agli utenti di cambiare le loro password.
- Crea identità individuali invece di credenziali condivise. Creando identità individuali, puoi assegnare a ciascun utente un set unico di credenziali di sicurezza. I singoli utenti consentono di sottoporre ad audit l'attività di ciascuno.

IAMRaccomandazioni dell'Identity Center:

- IAMIdentity Center fornisce una politica di password predefinita quando si utilizza la directory predefinita che stabilisce i requisiti di lunghezza, complessità e riutilizzo della password.
- Attiva MFA e configura l'impostazione sensibile al contesto o sempre attiva per MFA quando l'origine dell'identità è la directory predefinita o AD Connector AWS Managed Microsoft AD.
- Consenti agli utenti di registrare i propri dispositivi. MFA

Consigli sulle directory dei pool di utenti Amazon Cognito:

- Configura le impostazioni relative alla complessità della password.
- Richiesto MFA per gli utenti.
- Le impostazioni di sicurezza avanzate dei pool di utenti di Amazon Cognito offrono funzionalità come l'autenticazione adattiva, che può bloccare gli accessi sospetti.

IAMconsigli per gli utenti:

- Idealmente si utilizza IAM Identity Center o la federazione diretta. Tuttavia, potresti avere bisogno di IAM utenti. In tal caso, imposta una politica di password per IAM gli utenti. Puoi utilizzare la policy sulla password per definire requisiti quali la lunghezza minima o la necessità che la password richieda caratteri non alfabetici.
- Crea una IAM politica per imporre MFA l'accesso in modo che gli utenti possano gestire le proprie password e i propri dispositivi. MFA

Risorse

Best practice correlate:

- SEC02-BP03 Archivia e utilizza i segreti in modo sicuro

- SEC02-BP04 Affidati a un provider di identità centralizzato
- SEC03-BP08 Condividi le risorse in modo sicuro all'interno della tua organizzazione

Documenti correlati:

- AWS IAM Identity Center Politica in materia di password
- IAMpolitica sulle password degli utenti
- Impostazione della password dell'utente Account AWS root
- Policy sulla password Amazon Cognito
- AWS credenziali
- IAMmigliori pratiche di sicurezza

Video correlati:

- Gestione delle autorizzazioni degli utenti su larga scala con AWS IAM Identity Center
- Mastering identity at every layer of the cake

SEC02-BP02 Usa credenziali temporanee

Quando si esegue qualsiasi tipo di autenticazione, è preferibile utilizzare credenziali temporanee invece di credenziali a lungo termine per ridurre o eliminare i rischi, come la divulgazione, la condivisione o il furto involontario delle stesse.

Risultato desiderato: al fine di ridurre il rischio di credenziali a lungo termine, utilizza credenziali temporanee laddove possibile per le identità di persone e macchine. Le credenziali a lungo termine creano molti rischi, ad esempio possono essere caricate in forma di codice in archivi pubblici. GitHub Grazie alle credenziali temporanee, riduci notevolmente le possibilità di compromissione delle credenziali.

Anti-pattern comuni:

- Gli sviluppatori utilizzano chiavi di accesso a lungo termine fornite dagli IAM utenti anziché ottenere credenziali temporanee dalla federazione che li utilizza. CLI
- Sviluppatori che inseriscono chiavi di accesso a lungo termine nel loro codice e caricano tale codice su repository Git pubblici.

- Sviluppatori che inseriscono chiavi di accesso a lungo termine nelle app mobili che vengono poi rese disponibili negli app store.
- Utenti che condividono le chiavi di accesso a lungo termine con altri utenti o dipendenti che lasciano l'azienda con chiavi di accesso a lungo termine ancora in loro possesso.
- Utilizzo di chiavi di accesso a lungo termine per le identità macchina quando è possibile utilizzare credenziali temporanee.

Livello di rischio associato se questa best practice non fosse adottata: elevato

Guida all'implementazione

Utilizzate credenziali di sicurezza temporanee anziché credenziali a lungo termine per tutte le AWS API richieste. CLI APIe CLI le richieste ai AWS servizi devono, in quasi tutti i casi, essere firmate utilizzando chiavi di AWS accesso. Queste richieste possono essere firmate con credenziali temporanee o a lungo termine. L'unico caso in cui è necessario utilizzare credenziali a lungo termine, note anche come chiavi di accesso a lungo termine, è se si utilizza un IAMutente o l'utente Account AWS root. Quando si esegue la federazione AWS o si assume un IAMruolo tramite altri metodi, vengono generate credenziali temporanee. Anche quando si accede AWS Management Console utilizzando le credenziali di accesso, vengono generate credenziali temporanee per effettuare chiamate ai servizi. AWS Sono poche le situazioni in cui occorrono credenziali a lungo termine ed è possibile svolgere quasi tutte le attività utilizzando credenziali temporanee.

Evitare l'uso di credenziali a lungo termine a favore di credenziali temporanee dovrebbe andare di pari passo con una strategia di riduzione dell'utilizzo degli IAM utenti a favore della federazione e dei ruoli. IAM Sebbene in passato IAM gli utenti fossero utilizzati sia per l'identità umana che per quella automatica, ora consigliamo di non utilizzarli per evitare i rischi derivanti dall'uso di chiavi di accesso a lungo termine.

Passaggi dell'implementazione

Per le identità umane come dipendenti, amministratori, sviluppatori, operatori e clienti:

- È necessario affidarsi a un provider di identità centralizzato e richiedere agli utenti umani di utilizzare la federazione con un provider di identità per accedere AWS utilizzando credenziali temporanee. È possibile effettuare la federazione dei tuoi utenti con la federazione diretta di ciascun Account AWS o mediante AWS IAM Identity Center e il gestore dell'identità digitale di tua scelta. La federazione offre una serie di vantaggi rispetto all'utilizzo degli IAM utenti, oltre all'eliminazione delle credenziali a lungo termine. Gli utenti possono inoltre richiedere credenziali

temporanee dalla riga di comando per la federazione diretta o utilizzando IAMIdentity Center. Ciò significa che sono pochi i casi d'uso che richiedono IAM utenti o credenziali a lungo termine per gli utenti.

- Quando concedi a terzi, come i fornitori di software as a service (SaaS), l'accesso alle risorse del Account AWS tuo account, puoi utilizzare ruoli tra account e politiche basate sulle risorse.
- Se devi concedere alle applicazioni per consumatori o clienti l'accesso alle tue AWS risorse, puoi utilizzare i pool di identità di Amazon Cognito o i pool di utenti Amazon Cognito per fornire credenziali temporanee. Le autorizzazioni per le credenziali sono configurate tramite ruoli. IAM È inoltre possibile definire un IAM ruolo separato con autorizzazioni limitate per gli utenti ospiti che non sono autenticati.

Per le identità macchina, potrebbero essere necessarie credenziali a lungo termine. In questi casi, è necessario richiedere ai carichi di lavoro di utilizzare credenziali temporanee con IAM ruoli a cui accedere. AWS

- Per Amazon Elastic Compute Cloud (AmazonEC2), puoi utilizzare i ruoli per Amazon EC2.

- AWS Lambdaconsente di configurare un ruolo di esecuzione Lambda per concedere al servizio le autorizzazioni per eseguire AWS azioni utilizzando credenziali temporanee. Esistono molti altri modelli simili di AWS servizi per concedere credenziali temporanee utilizzando i ruoli. IAM
- Per i dispositivi IoT, puoi richiedere credenziali temporanee al provider di credenziali AWS IoT Core.
- Per i sistemi locali o i sistemi eseguiti all'esterno AWS che richiedono l'accesso alle AWS risorse, puoi utilizzare IAMRoles Anywhere.

Esistono scenari in cui le credenziali temporanee non sono un'opzione e potrebbe essere necessario utilizzare credenziali a lungo termine. In queste situazioni, procedi con l'audit e ruota periodicamente le credenziali, oltre a ruotare con regolarità le chiavi di accesso per casi d'uso che richiedono credenziali a lungo termine. Alcuni esempi che potrebbero richiedere credenziali a lungo termine includono WordPress plugin e client di terze parti. AWS In situazioni in cui è necessario utilizzare credenziali a lungo termine o per credenziali diverse dalle chiavi di AWS accesso, come gli accessi al database, è possibile utilizzare un servizio progettato per gestire la gestione dei segreti, ad esempio. AWS Secrets Manager Secrets Manager semplifica la gestione, la rotazione e lo storage sicuro delle chiavi segrete crittografate utilizzando i servizi supportati. Per ulteriori informazioni sulla rotazione delle credenziali a lungo termine, consulta rotazione delle chiavi di accesso.

Risorse

Best practice correlate:

- SEC02-BP03 Archivia e utilizza i segreti in modo sicuro
- SEC02-BP04 Affidati a un provider di identità centralizzato
- SEC03-BP08 Condividi le risorse in modo sicuro all'interno della tua organizzazione

Documenti correlati:

- Temporary Security Credentials
- Credenziali AWS
- Best practice sulla sicurezza IAM
- IAMRuoli
- IAMCentro di identità
- Identity Providers and Federation
- Rotating Access Keys
- Soluzioni dei partner per la sicurezza: accesso e controllo degli accessi
- L'utente root dell'account AWS

Video correlati:

- Gestione delle autorizzazioni degli utenti su larga scala con AWS IAM Identity Center
- Mastering identity at every layer of the cake

SEC02-BP03 Archivia e utilizza i segreti in modo sicuro

Un carico di lavoro richiede una capacità automatizzata di dimostrare la propria identità a database, risorse e servizi di terze parti. Ciò si ottiene utilizzando credenziali di accesso segrete, come chiavi di API accesso, password e token. OAuth L'utilizzo di un servizio appositamente creato per archiviare, gestire e ruotare queste credenziali aiuta a ridurre la probabilità che queste vengano compromesse.

Risultato desiderato: implementazione di un meccanismo per la gestione sicura delle credenziali delle applicazioni che consenta di raggiungere i seguenti obiettivi.

- Identificare i segreti necessari per il carico di lavoro.

- Ridurre il numero di credenziali a lungo termine sostituendole con credenziali a breve termine, laddove possibile.
- Stabilire l'archiviazione sicura e la rotazione automatica delle rimanenti credenziali a lungo termine.
- Sottoporre a audit l'accesso ai segreti esistenti nel carico di lavoro.
- Eseguire il monitoraggio continuo per verificare che nessun segreto sia incorporato nel codice sorgente durante il processo di sviluppo.
- Ridurre la probabilità che le credenziali vengano divulgate inavvertitamente.

Anti-pattern comuni:

- Nessuna rotazione delle credenziali.
- Memorizzazione di credenziali a lungo termine nel codice sorgente o nei file di configurazione.
- Memorizzazione delle credenziali a riposo non criptate.

Vantaggi dell'adozione di questa best practice:

- I segreti sono conservati in modo criptato a riposo e in transito.
- L'accesso alle credenziali è controllato tramite un API (pensalo come un distributore automatico di credenziali).
- L'accesso alle credenziali (sia in lettura che in scrittura) viene sottoposto a audit e registrato.
- Separazione delle preoccupazioni: la rotazione delle credenziali viene eseguita da un componente distinto, che può essere segregato dal resto dell'architettura.
- La distribuzione dei segreti avviene in automatico on demand ai componenti software e la rotazione avviene in una posizione centrale.
- È possibile controllare l'accesso alle credenziali in modo granulare.

Livello di rischio associato se questa best practice non fosse adottata: elevato

Guida all'implementazione

In passato, le credenziali utilizzate per l'autenticazione su databaseAPIs, token e altri segreti di terze parti potevano essere incorporate nel codice sorgente o nei file di ambiente. AWS offre diversi meccanismi per archiviare queste credenziali in modo sicuro, ruotarle automaticamente e controllarne l'utilizzo.

Il modo migliore per affrontare la gestione dei segreti è seguire le indicazioni relative a rimozione, sostituzione e rotazione. Le credenziali più sicure sono quelle che non si devono memorizzare, gestire o trattare. Possono esserci credenziali non più necessarie per il funzionamento del carico di lavoro e che possono essere rimosse in modo sicuro.

Per le credenziali ancora necessarie per il corretto funzionamento del carico di lavoro, potrebbe esserci l'opportunità di sostituire le credenziali a lungo termine con credenziali temporanee o a breve termine. Ad esempio, anziché codificare una chiave di accesso AWS segreta, prendi in considerazione la possibilità di sostituire quella credenziale a lungo termine con una credenziale temporanea che utilizza i ruoli. IAM

Alcuni segreti di lunga durata potrebbero non poter essere rimossi o sostituiti. È possibile archiviare tali segreti in un servizio come AWS Secrets Manager, dove saranno archiviati, gestiti e rotati a livello centrale su base regolare.

Un audit del codice sorgente e dei file di configurazione del carico di lavoro può rivelare molti tipi di credenziali. La tabella seguente riassume le strategie per gestire i tipi più comuni di credenziali:

Tipo di credenziali	Descrizione	Strategia suggerita
IAMchiavi di accesso	AWS IAMchiavi di accesso e segrete utilizzate per assumere IAM ruoli all'interno di un carico di lavoro	Sostituisci: utilizza invece IAMi ruoli assegnati alle istanze di calcolo (come Amazon EC2 o AWS Lambda). Per l'interoperabilità con terze parti che richiedono l'accesso alle risorse del tuo account Account AWS, chiedi se supportano l'accesso su più account.AWS Per le app mobili, prendi in considera zione l'utilizzo di credenziali temporanee tramite pool di identità di Amazon Cognito (identità federate). Per i carichi di lavoro eseguiti al di fuori di AWS, prendi in considerazione IAMRoles Anywhere o AWS

Tipo di credenziali	Descrizione	Strategia suggerita
		Systems Manager Hybrid Activations.
SSHchiavi	chiavi private Secure Shell utilizzate per accedere alle EC2 istanze Linux, manualmente o come parte di un processo automatizzato	Sostituisci: utilizza AWS Systems Manager o EC2Instance Connect per fornire accesso programma tico e umano alle EC2 istanze utilizzando IAM i ruoli.
Credenziali di applicazione e database	Password: stringa di testo semplice	Rotazione: memorizza le credenziali in AWS Secrets Manager e, laddove possibile , stabilisci una rotazione automatica.
Credenziali del database di amministrazione di Amazon RDS e Aurora	Password: stringa di testo semplice	Sostituisci: utilizza l'integrazi one di Secrets Manager con Amazon RDS o Amazon Aurora. Inoltre, alcuni tipi di RDS database possono utilizzare IAM ruoli anziché password per alcuni casi d'uso (per maggiori dettagli, consulta l'autenticazione del IAM database).
OAuthgettoni	Token segreti: stringa di testo semplice	Rotazione: archivia i token in AWS Secrets Manager e configura la rotazione automatica.
APIgettoni e chiavi	Token segreti: stringa di testo semplice	Rotazione: archivia in AWS Secrets Manager e stabilisc i una rotazione automatica, laddove possibile.

Un anti-pattern comune consiste nell'incorporare le chiavi di IAM accesso all'interno del codice sorgente, dei file di configurazione o delle app mobili. Quando è necessaria una chiave di IAM accesso per comunicare con un AWS servizio, utilizzate credenziali di sicurezza temporanee (a breve termine). Queste credenziali a breve termine possono essere fornite tramite IAMruoli per EC2 le istanze, ruoli di esecuzione per le funzioni Lambda, ruoli Cognito per l'accesso degli utenti mobili e policy IoT Core per IAM i dispositivi IoT. Quando ti interfacci con terze parti, preferisci delegare l'accesso a un IAM ruolo con l'accesso necessario alle risorse del tuo account piuttosto che configurare un IAM utente e inviare alla terza parte la chiave di accesso segreta per quell'utente.

Esistono molti casi in cui il carico di lavoro richiede l'archiviazione dei segreti necessari per interagire con altri servizi e risorse. AWS Secrets Managerè stato creato appositamente per gestire in modo sicuro queste credenziali, nonché l'archiviazione, l'uso e la rotazione di API token, password e altre credenziali.

AWS Secrets Manager offre cinque funzionalità chiave per garantire l'archiviazione e la gestione sicure delle credenziali sensibili: crittografia inattiva, crittografia intransito, controllo completo, controllo granulare degli accessi e rotazione estensibile dellecredenziali. Sono accettabili anche altri servizi di gestione dei segreti dei partner AWS o soluzioni sviluppate localmente che forniscano funzionalità e garanzie simili.

Passaggi dell'implementazione

1. Identifica i percorsi di codice contenenti credenziali codificate utilizzando strumenti automatizzati come Amazon. CodeGuru
   a. Usa Amazon CodeGuru per scansionare i tuoi repository di codice. Una volta completata la revisione, filtra CodeGuru per trovare righe di codice problematiche. `Type=Secrets`
2. Identifica le credenziali che possono essere rimosse o sostituite.
   a. Identifica le credenziali non più necessarie e contrassegnarle per la rimozione.
   b. Per le chiavi AWS segrete incorporate nel codice sorgente, sostituiscile con IAM ruoli associati alle risorse necessarie. Se parte del carico di lavoro è esterno AWS ma richiede IAM credenziali per accedere alle AWS risorse, prendi in considerazione IAMRoles Anywhere o AWS Systems Manager Hybrid Activations.
3. Per altri segreti di terze parti a lunga durata che richiedono l'uso della strategia di rotazione, integra Secrets Manager nel codice per recuperare i segreti di terze parti in fase di esecuzione.
   a. La CodeGuru console può creare automaticamente un segreto in Secrets Manager utilizzando le credenziali scoperte.
   b. Integra il recupero dei segreti da Secrets Manager nel codice dell'applicazione.

   i. Le funzioni Lambda serverless possono utilizzare un'estensione Lambda indipendente dal linguaggio.
   ii. Per EC2 istanze o contenitori, AWS fornisce un esempio di codice lato client per il recupero di segreti da Secrets Manager in diversi linguaggi di programmazione popolari.
4. Esamina periodicamente la base di codice e ripetere la scansione per verificare che non siano stati aggiunti nuovi segreti al codice.
   a. Prendi in considerazione l'utilizzo di uno strumento come git-secrets per evitare di inserire nuovi segreti nel tuo repository di codice sorgente.
5. Monitora l'attività di Secrets Manager per individuare eventuali indicazioni di utilizzo imprevisto, accesso inopportuno ai segreti o tentativi di eliminazione degli stessi.
6. Riduci l'esposizione umana alle credenziali. Limita l'accesso alla lettura, alla scrittura e alla modifica delle credenziali a un IAM ruolo dedicato a questo scopo e consenti l'accesso per assumere il ruolo solo a un piccolo sottoinsieme di utenti operativi.

Risorse

Best practice correlate:

- SEC02-BP02 Usa credenziali temporanee
- SEC02-BP05 Verifica e ruota periodicamente le credenziali

Documenti correlati:

- Guida introduttiva con AWS Secrets Manager
- Identity Providers and Federation
- Amazon CodeGuru presenta Secrets Detector
- Come si usa AWS Secrets ManagerAWS Key Management Service
- Secret encryption and decryption in Secrets Manager
- Articoli del blog su Secrets Manager
- Amazon RDS annuncia l'integrazione con AWS Secrets Manager

Video correlati:

- Best Practices for Managing, Retrieving, and Rotating Secrets at Scale

- Trova segreti codificati con Amazon Secrets Detector CodeGuru
- Protezione dei segreti per i carichi di lavoro ibridi utilizzando AWS Secrets Manager

Workshop correlati:

- Archivia, recupera e gestisci le credenziali sensibili in AWS Secrets Manager
- AWS Systems Manager Attivazioni ibride

SEC02-BP04 Affidati a un provider di identità centralizzato

Per le identità della forza lavoro (dipendenti e collaboratori) affidati a un gestore dell'identità digitale che ti consenta di gestire le identità in un luogo centralizzato. In questo modo è più semplice gestire l'accesso tra più applicazioni e sistemi, poiché crei, assegni, gestisci, revochi e verifichi gli accessi da una singola posizione.

Risultato desiderato: disponi di un provider di identità centralizzato in cui gestire centralmente gli utenti della forza lavoro, le politiche di autenticazione (ad esempio la richiesta dell'autenticazione a più fattori (MFA)) e l'autorizzazione a sistemi e applicazioni (come l'assegnazione dell'accesso in base all'appartenenza o agli attributi di un utente). Gli utenti che fanno parte della tua forza lavoro accedono al gestore dell'identità digitale centrale ed effettuano l'accesso federato (autenticazione unica) alle applicazioni interne ed esterne, il che elimina la necessità per gli utenti di ricordare più credenziali. Il gestore dell'identità digitale è integrato con i tuoi sistemi di risorse umane (HR), in modo che le modifiche relative al personale vengano sincronizzate in automatico con il gestore dell'identità digitale. Ad esempio, se qualcuno lascia l'organizzazione, puoi revocare automaticamente l'accesso ad applicazioni e sistemi federati (inclusi). AWS Hai abilitato la verifica dettagliata dei log nel tuo gestore dell'identità digitale e stai monitorando questi log per rilevare comportamenti degli utenti insoliti.

Anti-pattern comuni:

- Non utilizzi federazione e autenticazione unica. Gli utenti che appartengono alla tua forza lavoro creano account utente e credenziali separati in più applicazioni e sistemi.
- Non hai automatizzato il ciclo di vita delle identità degli utenti che fanno parte della tua forza lavoro, ad esempio integrando il gestore dell'identità digitale con i tuoi sistemi HR. Quando un utente lascia l'organizzazione o cambia ruolo, segui una procedura manuale per eliminare o aggiornare i suoi record in più applicazioni e sistemi.

Vantaggi dell'adozione di questa best practice: utilizzare un gestore dell'identità digitale centralizzato ti fornisce un'unica piattaforma per gestire le identità e le policy degli utenti che fanno parte della tua forza lavoro, la possibilità di assegnare l'accesso alle applicazioni a utenti e gruppi e di monitorare l'attività di accesso degli utenti. Grazie all'integrazione con i sistemi di risorse umane (HR), quando un utente cambia ruolo, queste modifiche vengono sincronizzate con il gestore dell'identità digitale e le applicazioni e le autorizzazioni assegnate si aggiornano in automatico. Quando un utente lascia l'organizzazione, la sua identità viene automaticamente disabilitata nel gestore dell'identità digitale e l'accesso alle applicazioni e ai sistemi federati viene revocato.

Livello di rischio associato se questa best practice non fosse adottata: elevato

Guida all'implementazione

Linee guida per l'accesso ad AWS degli utenti che fanno parte della forza lavoro

Gli utenti della forza lavoro, come dipendenti e collaboratori dell'organizzazione, potrebbero dover accedere all' AWS utilizzo di AWS Management Console or AWS Command Line Interface (AWS CLI) per svolgere le proprie mansioni lavorative. Puoi concedere AWS l'accesso agli utenti della tua forza lavoro federando il tuo provider di identità centralizzato a due livelli: federazione diretta AWS a ciascuno Account AWS o federazione a più account dell'organizzazione.AWS

- Per federare gli utenti della forza lavoro direttamente con ciascuno di essi Account AWS, puoi utilizzare un provider di identità centralizzato con cui eseguire la federazione in quell'account. AWS Identity and Access Management La flessibilità di IAM consente di abilitare un Identity Provider SAML2.0 o Open ID Connect (OIDC) separato per ciascuno Account AWS e di utilizzare attributi utente federati per il controllo degli accessi. Gli utenti della tua forza lavoro utilizzeranno il proprio browser web per accedere al provider di identità fornendo le proprie credenziali (come password e MFA codici token). Il provider di identità invia al proprio browser un'SAMLasserzione che viene inviata all'accesso per consentire URL all'utente di AWS Management Console accedere in modalità single sign-on assumendo un ruolo.AWS Management Console IAM Gli utenti possono inoltre ottenere AWS API credenziali temporanee da utilizzare in AWS CLIo AWS SDKsda AWS STSassumendo il IAM ruolo utilizzando un'SAMLasserzione del provider di identità.
- Per federare gli utenti della forza lavoro con più account all' AWS interno dell'organizzazione, è possibile gestire centralmente l'AWS IAM Identity Centeraccesso degli utenti della forza lavoro alle applicazioni e agli utenti. Account AWS Puoi abilitare il Centro identità per la tua organizzazione e configurare la tua origine di identità. IAMIdentity Center fornisce una directory di origine delle identità predefinita che puoi utilizzare per gestire utenti e gruppi. In alternativa, puoi scegliere un'origine di identità esterna connettendoti al tuo provider di identità esterno tramite la SAML

versione 2.0 e assegnando automaticamente il provisioning a utenti e gruppi utilizzando SCIM o connettendoti al tuo Microsoft AD Directory utilizzando AWS Directory Service. Una volta configurata un'origine di identità, è possibile assegnare l'accesso a utenti e gruppi Account AWS definendo politiche con privilegi minimi nei set di autorizzazioni. Gli utenti della tua forza lavoro possono autenticarsi tramite il tuo gestore dell'identità digitale centrale per accedere al portale di accesso AWS ed eseguire l'accesso tramite autenticazione unica per gli Account AWS e le applicazioni cloud a loro assegnate. Gli utenti possono configurare la versione AWS CLI v2 per l'autenticazione con Identity Center e ottenere le credenziali per eseguire i comandi. AWS CLI Identity Center consente inoltre l'accesso Single Sign-On ad AWS applicazioni come i SageMaker portali Amazon Studio e AWS IoT Sitewise Monitor.

Dopo aver seguito le indicazioni precedenti, gli utenti della forza lavoro non avranno più bisogno di utilizzare IAM utenti e gruppi per le normali operazioni durante la gestione dei carichi di lavoro su. AWSGli utenti e i gruppi vengono invece gestiti all'esterno AWS e gli utenti possono accedere alle AWS risorse come identità federata. Le identità federate utilizzano i gruppi definiti dal gestore dell'identità digitale centralizzato. È necessario identificare e rimuovere IAM gruppi, IAM utenti e credenziali utente di lunga durata (password e chiavi di accesso) che non sono più necessarie nel proprio. Account AWSPuoi trovare le credenziali inutilizzate utilizzando i report sulle IAM credenziali, eliminare gli utenti corrispondenti ed eliminare i gruppi. IAM IAM È possibile applicare una Service Control Policy (SCP) all'organizzazione che aiuta a prevenire la creazione di nuovi IAM utenti e gruppi, imponendo che l'accesso avvenga AWS tramite identità federate.

Linee guida per gli utenti delle tue applicazioni

Puoi gestire le identità degli utenti delle tue applicazioni, ad esempio un'app mobile, utilizzando Amazon Cognito come gestore dell'identità digitale centralizzato. Amazon Cognito consente l'autenticazione, autorizzazione e gestione degli utenti per le app Web e per dispositivi mobili. Amazon Cognito offre un archivio di identità scalabile fino a milioni di utenti, supporta la federazione delle identità sociali e aziendali e offre funzionalità di sicurezza avanzate per proteggere i tuoi utenti e la tua azienda. Puoi integrare la tua applicazione Web o mobile personalizzata con Amazon Cognito per aggiungere l'autenticazione degli utenti e il controllo degli accessi alle applicazioni in pochi minuti. Basato su standard di identità aperti come SAML Open ID Connect (OIDC), Amazon Cognito supporta diverse normative di conformità e si integra con le risorse di sviluppo frontend e backend.

Passaggi dell'implementazione

Passaggi per l'accesso ad AWS degli utenti della forza lavoro

- Concedi agli utenti della tua forza lavoro di utilizzare un provider di identità centralizzato AWS utilizzando uno dei seguenti approcci:
  - Usa IAM Identity Center per abilitare il Single Sign-On a più Account AWS utenti della tua AWS organizzazione tramite la federazione con il tuo provider di identità.
  - Utilizzalo IAM per connettere il tuo provider di identità direttamente a ciascuno di essi Account AWS, abilitando un accesso federato e granulare.
- Identifica e rimuovi IAM utenti e gruppi che vengono sostituiti da identità federate.

Passaggi per gli utenti delle tue applicazioni

- Utilizza Amazon Cognito come gestore dell'identità digitale centralizzato per le tue applicazioni.
- Integra le tue applicazioni personalizzate con Amazon Cognito utilizzando OpenID Connect e. OAuth Puoi sviluppare le tue applicazioni personalizzate utilizzando le librerie Amplify che forniscono interfacce semplici da integrare con una varietà AWS di servizi, come Amazon Cognito per l'autenticazione.

Risorse

Best practice Well-Architected correlate:

- SEC02-BP06 Utilizza gruppi e attributi di utenti
- SEC03-BP02 Concedi l'accesso con privilegi minimi
- SEC03-BP06 Gestisci l'accesso in base al ciclo di vita

Documenti correlati:

- Federazione delle identità in AWS
- Best practice relative alla sicurezza in IAM
- AWS Identity and Access Management Best practice
- Guida introduttiva all'IAMamministrazione delegata di Identity Center
- Come utilizzare le policy gestite dai clienti in IAM Identity Center per casi d'uso avanzati
- AWS CLI v2: fornitore di credenziali IAM Identity Center

Video correlati:

- AWS re:Inforce 2022 - () approfondimento AWS Identity and Access Management IAM
- AWS re:Invent 2022 - Semplifica l'accesso alla forza lavoro esistente con Identity Center IAM
- AWS re:Invent 2018: padroneggiare l'identità a ogni livello

Esempi correlati:

- Workshop: Utilizzo AWS IAM Identity Center per ottenere una solida gestione dell'identità
- Workshop: Serverless identity

Strumenti correlati:

- AWS Partner con competenze in materia di sicurezza: Identity and Access Management
- AWS IAM Identity Center

SEC02-BP05 Verifica e ruota periodicamente le credenziali

Sottoponi a audit e ruota periodicamente le credenziali per limitarne il tempo di utilizzo per l'accesso alle risorse. Le credenziali a lungo termine espongono a molti rischi, riducibili mediante la rotazione periodica.

Risultato desiderato: implementa la rotazione delle credenziali per ridurre i rischi associati all'utilizzo delle credenziali a lungo termine. Esegui regolarmente l'audit e rimedia alla non conformità con le policy di rotazione delle credenziali.

Anti-pattern comuni:

- Nessun audit dell'uso delle credenziali.
- Utilizzo non necessario di credenziali a lungo termine.
- Utilizzo di credenziali a lungo termine e mancata rotazione regolare.

Livello di rischio associato se questa best practice non fosse adottata: elevato

Guida all'implementazione

Quando non puoi fare affidamento su credenziali temporanee e richiedere credenziali a lungo termine, verifica le credenziali per verificare che controlli definiti come l'autenticazione a più fattori (MFA) vengano applicati, ruotati regolarmente e abbiano il livello di accesso appropriato.

La convalida periodica, preferibilmente tramite uno strumento automatizzato, è necessaria per verificare l'applicazione dei controlli corretti. Per le identità umane, è necessario richiedere agli utenti di modificare periodicamente le password e ritirare le chiavi di accesso a favore delle credenziali temporanee. Passando da AWS Identity and Access Management (IAM) utenti a identità centralizzate, potete generare un rapporto sulle credenziali per controllare gli utenti.

Ti consigliamo inoltre di applicare e monitorare MFA il tuo provider di identità. Puoi impostare Regole di AWS Configo utilizzare gli standard di AWS Security Hub sicurezza per monitorare se gli utenti hanno configurato MFA la configurazione. Prendi in considerazione l'utilizzo di IAM Roles Anywhere per fornire credenziali temporanee per le identità delle macchine. In situazioni in cui non è possibile utilizzare IAM ruoli e credenziali temporanee, è necessario controllare frequentemente e ruotare le chiavi di accesso.

Passaggi dell'implementazione

- Controlla regolarmente le credenziali: verifica le identità configurate nel tuo provider di identità e IAM aiuta a verificare che solo le identità autorizzate abbiano accesso al tuo carico di lavoro. Tali identità possono includere, a titolo esemplificativo ma non esaustivo, IAM utenti, AWS IAM Identity Center utenti di Active Directory o utenti di un provider di identità upstream diverso. Ad esempio, eliminare le persone che lasciano l'organizzazione e i ruoli multi-account non più necessari. Predisponi di una procedura per verificare periodicamente le autorizzazioni ai servizi a cui accede un'entità. IAM In questo modo potrai identificare le policy da modificare per rimuovere le autorizzazioni non utilizzate. Utilizzate i report sulle credenziali e AWS Identity and Access Management Access Analyzercontrollate IAM le credenziali e le autorizzazioni. Puoi usare Amazon CloudWatch per impostare allarmi per API chiamate specifiche chiamate all'interno del tuo AWS ambiente. Amazon GuardDuty può anche avvisarti di attività impreviste, che potrebbero indicare un accesso eccessivamente permissivo o un accesso non intenzionale alle credenziali. IAM
- Ruota le credenziali regolarmente: quando non puoi utilizzare credenziali temporanee, ruota le chiavi di IAM accesso a lungo termine regolarmente (al massimo ogni 90 giorni). In caso di divulgazione involontaria e a propria insaputa di una chiave di accesso, questo limita la durata di utilizzo delle credenziali per accedere alle risorse. Per informazioni sulla rotazione delle chiavi di accesso per IAM gli utenti, consulta Rotazione delle chiavi di accesso.
- Rivedi IAM le autorizzazioni: per migliorare la sicurezza delle tue politiche Account AWS, rivedi e monitora regolarmente ciascuna delle tue politiche. IAM Verifica che le policy rispettino il principio del privilegio minimo.
- Valuta la possibilità di automatizzare la creazione e gli aggiornamenti IAM delle risorse: IAM Identity Center automatizza molte IAM attività, come la gestione di ruoli e policy. In alternativa,

AWS CloudFormation può essere utilizzato per automatizzare l'implementazione delle IAM risorse, inclusi ruoli e policy, per ridurre la possibilità di errori umani, poiché i modelli possono essere verificati e le versioni controllate.

- Usa IAM Roles Anywhere per sostituire IAM gli utenti per le identità delle macchine: IAM Roles Anywhere ti consente di utilizzare i ruoli in aree che tradizionalmente non consentivi, come i server locali. IAMRoles Anywhere utilizza un certificato X.509 affidabile per autenticarsi e ricevere credenziali temporanee. AWS L'utilizzo di IAM Roles Anywhere evita la necessità di ruotare queste credenziali, poiché le credenziali a lungo termine non vengono più archiviate nell'ambiente locale. È necessario monitorare e ruotare il certificato X.509 quando si avvicina alla scadenza.

Risorse

Best practice correlate:

- SEC02-BP02 Usa credenziali temporanee
- SEC02-BP03 Archivia e utilizza i segreti in modo sicuro

Documenti correlati:

- Guida introduttiva con AWS Secrets Manager
- IAMMigliori pratiche
- Identity Providers and Federation
- Soluzioni dei partner per la sicurezza: accesso e controllo degli accessi
- Temporary Security Credentials
- Ottenere report sulle credenziali per il tuo Account AWS

Video correlati:

- Best Practices for Managing, Retrieving, and Rotating Secrets at Scale
- Gestione delle autorizzazioni degli utenti su larga scala con AWS IAM Identity Center
- Mastering identity at every layer of the cake

Esempi correlati:

- Well-Architected Lab - Pulizia automatica degli utenti IAM

- Well-Architected Lab - Distribuzione IAM automatizzata di gruppi e ruoli

SEC02-BP06 Utilizza gruppi e attributi di utenti

Definire le autorizzazioni in base a gruppi di utenti e attributi aiuta a ridurre numero e complessità delle policy, semplificando il raggiungimento del principio del privilegio minimo. Puoi usare i gruppi di utenti per gestire le autorizzazioni di molte persone in un'unica posizione, in base alla funzione svolte nell'organizzazione. Gli attributi, come il reparto o la sede, possono fornire un ulteriore livello di portata dei permessi quando le persone svolgono una funzione simile ma per sottoinsiemi diversi di risorse.

Risultato desiderato: puoi applicare modifiche alle autorizzazioni in base alla funzione per tutti gli utenti che la eseguono. L'appartenenza al gruppo e gli attributi regolano le autorizzazioni degli utenti, riducendo la necessità di gestire le autorizzazioni a livello di singolo utente. I gruppi e gli attributi definiti nel gestore dell'identità digitale vengono propagati automaticamente agli ambienti AWS .

Anti-pattern comuni:

- Gestione delle autorizzazioni per singoli utenti e duplicazione tra più utenti.
- Definizione dei gruppi a un livello troppo alto, concessione di autorizzazioni troppo estese.
- Definizione di gruppi a un livello troppo granulare, che crea duplicazioni e confusione sull'appartenenza.
- Utilizzo di gruppi con autorizzazioni duplicate su sottoinsiemi di risorse quando è possibile utilizzare invece gli attributi.
- Nessuna gestione di gruppi, attributi e appartenenze attraverso un gestore dell'identità digitale standardizzato e integrato con gli ambienti AWS .

Livello di rischio associato se questa best practice non fosse adottata: medio

Guida all'implementazione

AWS le autorizzazioni sono definite in documenti denominati policy associati a un principale, ad esempio un utente, un gruppo, un ruolo o una risorsa. Per la forza lavoro, ciò consente di definire i gruppi in base alla funzione svolta dagli utenti per l'organizzazione, anziché in base alle risorse a cui si accede. Ad esempio, un `WebAppDeveloper` gruppo può avere una politica allegata per la configurazione di un servizio come Amazon CloudFront all'interno di un account di sviluppo. Un `AutomationDeveloper` gruppo può avere alcune CloudFront autorizzazioni in comune con

il `WebAppDeveloper` gruppo. È possibile inserire queste autorizzazioni in una policy separata e associarle a entrambi i gruppi, anziché far appartenere a un gruppo `CloudFrontAccess` gli utenti di entrambe le funzioni.

Oltre ai gruppi, è possibile utilizzare gli attributi per un ulteriore ambito dell'accesso. Ad esempio, potresti avere un attributo `Project` che consente agli utenti del tuo gruppo `WebAppDeveloper` di stabilire l'accesso a risorse specifiche del loro progetto. L'uso di questa tecnica elimina la necessità di avere gruppi diversi per gli sviluppatori di applicazioni che lavorano su progetti diversi, se le loro autorizzazioni sono comunque le stesse. Il modo in cui si fa riferimento agli attributi nelle politiche di autorizzazione si basa sulla loro origine, indipendentemente dal fatto che siano definiti come parte del protocollo di federazione (ad esempioSAML, oSCIM)OIDC, come SAML asserzioni personalizzate o impostati all'interno di IAM Identity Center.

Passaggi dell'implementazione

1. Stabilisci dove definire gruppi e attributi.
   a. Seguendo le indicazioni riportate inSEC02-BP04 Affidati a un provider di identità centralizzato, è possibile determinare se è necessario definire gruppi e attributi all'interno del proprio provider di identità, all'interno di IAM Identity Center o utilizzare i gruppi di IAM utenti in un account specifico.
2. Definisci i gruppi.
   a. Determina i tuoi gruppi in base alla funzione e all'ambito di accesso richiesti.
   b. Se lo definisci all'interno di IAM Identity Center, crea gruppi e associa il livello di accesso desiderato utilizzando i set di autorizzazioni.
   c. Se lo definisci all'interno di un provider di identità esterno, stabilisci se il provider supporta il SCIM protocollo e valuta la possibilità di abilitare il provisioning automatico all'interno di IAM Identity Center. Questa funzionalità sincronizza la creazione, l'appartenenza e l'eliminazione dei gruppi tra il provider e IAM Identity Center.
3. Definizione degli attributi.
   a. Se si utilizza un provider di identità esterno, entrambi i protocolli SCIM e SAML 2.0 forniscono determinati attributi per impostazione predefinita. È possibile definire e passare attributi aggiuntivi utilizzando SAML asserzioni utilizzando il nome dell'`https://aws.amazon.com/SAML/Attributes/PrincipalTag`attributo.
   b. Se la definizione viene effettuata all'interno di IAM Identity Center, abilitate la funzionalità di controllo degli accessi basata sugli attributi (ABAC) e definite gli attributi desiderati.
4. Autorizzazioni di ambito basate su gruppi e attributi.

a. Prendi in considerazione la possibilità di includere nelle tue policy di autorizzazione condizioni che confrontino gli attributi del tuo principale con gli attributi delle risorse a cui si accede. Ad esempio, puoi definire una condizione che consenta l'accesso a una risorsa solo se il valore di una chiave di condizione `PrincipalTag` corrisponde a quello di una chiave `ResourceTag` con lo stesso nome.

Risorse

Best practice correlate:

- SEC02-BP04 Affidati a un provider di identità centralizzato
- SEC03-BP02 Concedi l'accesso con privilegi minimi
- COST02-BP04 Implementa gruppi e ruoli

Documenti correlati:

- IAMMigliori pratiche
- Gestisci le identità in IAM Identity Center
- A cosa ABAC serve AWS?
- ABACIn IAM Identity Center

Video correlati:

- Gestione delle autorizzazioni degli utenti su larga scala con AWS IAM Identity Center
- Mastering identity at every layer of the cake

SEC3. Come si gestiscono le autorizzazioni per persone e macchine?

Gestisci le autorizzazioni per controllare l'accesso alle identità di persone e macchine che richiedono l'accesso e il tuo carico di lavoro. AWS Le autorizzazioni controllano chi può accedere a cosa e a quali condizioni.

Best practice

- SEC03-BP01 Definire i requisiti di accesso
- SEC03-BP02 Concedi l'accesso con privilegi minimi

- SEC03-BP03 Stabilire un processo di accesso di emergenza
- SEC03-BP04 Ridurre continuamente le autorizzazioni
- SEC03-BP05 Definisci barriere di autorizzazione per la tua organizzazione
- SEC03-BP06 Gestisci l'accesso in base al ciclo di vita
- SEC03-BP07 Analizza l'accesso pubblico e tra account
- SEC03-BP08 Condividi le risorse in modo sicuro all'interno della tua organizzazione
- SEC03-BP09 Condividi le risorse in modo sicuro con terze parti

SEC03-BP01 Definire i requisiti di accesso

Ogni componente o risorsa del carico di lavoro deve essere accessibile da amministratori, utenti finali o altri componenti. Definisci chiaramente chi o cosa deve avere accesso a ciascun componente e scegli il tipo di identità e il metodo di autenticazione e autorizzazione appropriati.

Anti-pattern comuni:

- Codifica fissa o archiviazione dei segreti nell'applicazione.
- Concessione di autorizzazioni personalizzate per ogni utente.
- Utilizzo di credenziali di lunga durata.

Livello di rischio associato se questa best practice non fosse adottata: elevato

Guida all'implementazione

Ogni componente o risorsa del carico di lavoro deve essere accessibile da amministratori, utenti finali o altri componenti. Definisci chiaramente chi o cosa deve avere accesso a ciascun componente e scegli il tipo di identità e il metodo di autenticazione e autorizzazione appropriati.

L'accesso regolare all' Account AWS interno dell'organizzazione deve essere fornito utilizzando un accesso federato o un provider di identità centralizzato. È inoltre necessario centralizzare la gestione delle identità e garantire che esista una pratica consolidata per integrare AWS l'accesso al ciclo di vita degli accessi dei dipendenti. Ad esempio, se un dipendente passa a un ruolo lavorativo con un livello di accesso diverso, anche la sua appartenenza al gruppo deve cambiare per riflettere i nuovi requisiti di accesso.

Nel definire i requisiti di accesso per le identità non umane, determina quali applicazioni e componenti devono accedere, nonché le modalità di concessione delle autorizzazioni. L'utilizzo di IAM ruoli creati

con il modello di accesso con privilegi minimi è un approccio consigliato. AWS Le policy gestite forniscono policy predefinite IAM che coprono i casi d'uso più comuni.

AWS i servizi, come AWS Secrets ManagerAWSSystems Manager Parameter Store, possono aiutare a separare i segreti dall'applicazione o dal carico di lavoro in modo sicuro nei casi in cui non è possibile utilizzare i ruoli. IAM In Secrets Manager, puoi adottare la rotazione automatica delle credenziali. È possibile utilizzare Systems Manager per fare riferimento ai parametri negli script, nei comandi, nei SSM documenti, nei flussi di lavoro di configurazione e automazione utilizzando il nome univoco specificato al momento della creazione del parametro.

È possibile utilizzare AWS Identity and Access Management Roles Anywhere per ottenere credenziali di sicurezza temporanee IAM per carichi di lavoro eseguiti all'esterno di. AWS I tuoi carichi di lavoro possono utilizzare le stesse IAMpolitiche e gli stessi IAMruoli che usi con AWS le applicazioni per accedere alle risorse. AWS

Ove possibile, prediligi le credenziali temporanee a breve termine rispetto a quelle statiche a lungo termine. Per gli scenari in cui occorrono utenti con accesso programmatico e credenziali a lungo termine, usa le ultime informazioni usate per la chiave di accesso per la rotazione e la rimozione delle chiavi di accesso.

Gli utenti necessitano dell'accesso programmatico se desiderano interagire con l' AWS esterno di. AWS Management Console Il modo per concedere l'accesso programmatico dipende dal tipo di utente che accede. AWS

Per fornire agli utenti l'accesso programmatico, scegli una delle seguenti opzioni.

Quale utente necessita dell'accesso programmatico?	Per	Come
Identità della forza lavoro (Utenti gestiti in IAM Identity Center)	Utilizza credenziali temporane e per firmare le richieste programmatiche a AWS CLI, AWS SDKs, o. AWS APIs	Segui le istruzioni per l'interfa ccia che desideri utilizzare. • Per la AWS CLI, vedere Configurazione dell'uso AWS IAM Identity Center nella AWS CLI Guida per l'utente.AWS Command Line Interface

Quale utente necessita dell'accesso programmatico?	Per	Come
		• Per AWS SDKs gli strumenti e AWS APIs, consulta l'autenticazione di IAM Identity Center nella Guida di riferimento agli strumenti AWS SDKs e agli strumenti.
IAM	Utilizza credenziali temporane e per firmare le richieste programmatiche a AWS CLI, AWS SDKs, o. AWS APIs	Seguendo le istruzioni riportate in Utilizzo delle credenziali temporanee con le AWS risorse nella Guida per l'IAMutente.
IAM	(Non consigliato) Utilizza credenziali a lungo termine per firmare richieste programmatiche a AWS CLI,, AWS SDKs o. AWS APIs	Segui le istruzioni per l'interfa ccia che desideri utilizzare. • Per la AWS CLI, consulta Autenticazione tramite credenziali IAM utente nella Guida per l'utente.AWS Command Line Interface • Per AWS SDKs gli strumenti , consulta Autenticazione tramite credenziali a lungo termine nella Guida di riferimento agli strumenti e agli AWS SDKs strumenti. • Per AWS APIs, consulta Gestione delle chiavi di accesso per IAM gli utenti nella Guida per l'IAMutente.

Risorse

Documenti correlati:

- Controllo dell'accesso basato sugli attributi () ABAC
- AWS IAM Identity Center
- IAM Roles Anywhere
- AWS Politiche gestite per Identity Center IAM
- AWS IAMcondizioni politiche
- IAMcasi d'uso
- Rimuovere credenziali non necessarie
- Lavorare con le policy
- Come controllare l'accesso alle AWS risorse in base Account AWS all'unità organizzativa o all'organizzazione
- Identifica, organizza e gestisci facilmente i segreti utilizzando la ricerca avanzata in AWS Secrets Manager

Video correlati:

- Diventa un IAM policy master in 60 minuti o meno
- Separation of Duties, Least Privilege, Delegation, and CI/CD
- Streamlining identity and access management for innovation

SEC03-BP02 Concedi l'accesso con privilegi minimi

È una best practice concedere alle identità soltanto il livello di accesso di cui hanno bisogno, specificando le operazioni che possono effettuare, le risorse su cui possono operare e a quali condizioni. Affidati a gruppi e attributi di identità per impostare in modo dinamico le autorizzazioni su vasta scala, anziché definire le autorizzazioni per i singoli utenti. Ad esempio, puoi concedere a un gruppo di sviluppatori le autorizzazioni per gestire solo le risorse del loro progetto. In questo modo, se uno sviluppatore lascia il progetto, il suo accesso viene revocato in automatico senza modificare le policy di accesso sottostanti.

Risultato desiderato: gli utenti dispongono solo delle autorizzazioni necessarie per svolgere il proprio lavoro. Gli utenti dovrebbero avere accesso solo agli ambienti di produzione per eseguire un'attività specifica in un intervallo temporale limitato e l'accesso dovrebbe essere revocato

una volta completata l'attività. Le autorizzazioni devono essere revocate quando non sono più necessarie, incluso se un utente passa a un progetto o a un ruolo professionale diverso. I privilegi di amministratore devono essere riservati a un piccolo gruppo di amministratori fidati. Le autorizzazioni vanno riviste con regolarità per evitare che si accumulino. Account di sistemi o di macchine devono disporre del numero minimo di autorizzazioni necessarie per portare a termine un'attività.

Anti-pattern comuni:

- L'impostazione predefinita è la concessione delle autorizzazioni di amministratore agli utenti.
- Utilizzo dell'utente root per le attività. day-to-day
- Creazione di policy eccessivamente permissive, ma senza privilegi completi di amministratore.
- Mancata revisione delle autorizzazioni per capire se consentono l'accesso privilegio minimo.

Livello di rischio associato se questa best practice non fosse adottata: elevato

Guida all'implementazione

Secondo il principio del privilegio minimo, le identità dovrebbero essere autorizzate a eseguire solo il più piccolo insieme di azioni necessarie per lo svolgimento di un'attività specifica. In questo modo usabilità, efficienza e sicurezza sono bilanciate. Seguendo questo principio si limitano gli accessi indesiderati e si può monitorare chi accede a quali risorse. IAMper impostazione predefinita, gli utenti e i ruoli non dispongono di autorizzazioni. Per impostazione predefinita, l'utente root dispone dell'accesso completo e deve essere strettamente controllato, monitorato e utilizzato solo per le attività che richiedono l'accesso root.

IAMle politiche vengono utilizzate per concedere esplicitamente le autorizzazioni a IAM ruoli o risorse specifiche. Ad esempio, le policy basate sull'identità possono essere collegate ai IAM gruppi, mentre i bucket S3 possono essere controllati da policy basate sulle risorse.

Quando si crea una IAM policy, è possibile specificare le azioni, le risorse e le condizioni di servizio che devono essere soddisfatte per consentire o negare l'accesso. AWS AWS supporta una serie di condizioni per aiutarti a limitare l'accesso. Ad esempio, utilizzando la chiave `PrincipalOrgID` condition, puoi negare azioni se il richiedente non fa parte della tua AWS organizzazione.

Puoi anche controllare le richieste che AWS i servizi effettuano per tuo conto, come la AWS CloudFormation creazione di una AWS Lambda funzione, utilizzando il tasto `CalledVia` condition. È necessario sovrapporre diversi tipi di policy per stabilire defense-in-depth e limitare le autorizzazioni complessive degli utenti. Puoi anche limitare le autorizzazioni che possono essere concesse e le

relative condizioni. Ad esempio, potete consentire ai team addetti alle applicazioni di creare IAM le proprie policy per i sistemi da loro creati, ma dovete anche applicare un limite di autorizzazione per limitare il numero massimo di autorizzazioni che il sistema può ricevere.

Passaggi dell'implementazione

- Implementa politiche con privilegi minimi: assegna politiche di accesso con privilegi minimi a IAM gruppi e ruoli in modo che riflettano il ruolo o la funzione dell'utente che hai definito.
  - Politiche di base sull'APIutilizzo: un modo per determinare le autorizzazioni necessarie consiste nel rivedere i registri. AWS CloudTrail Questa revisione consente di creare autorizzazioni personalizzate in base alle azioni effettivamente eseguite dall'utente all'interno. AWSIAMAccess Analyzer può generare automaticamente una IAM policy basata sull'attività. È possibile utilizzare IAM Access Advisor a livello di organizzazione o account per tenere traccia delle ultime informazioni a cui si accede per una determinata politica.
- Prendi in considerazione l'utilizzo di policy AWS gestite per le funzioni lavorative. Quando si inizia a creare politiche di autorizzazione granulari, può essere difficile sapere da dove iniziare. AWS ha gestito politiche per ruoli lavorativi comuni, ad esempio fatturazione, amministratori di database e data scientist. Queste policy possono contribuire a limitare l'accesso degli utenti e, al contempo, definiscono come implementare le policy di privilegio minimo.
- Rimuovi le autorizzazioni non necessarie: rimuovi le autorizzazioni non necessarie e riduci le policy eccessivamente permissive. IAMLa generazione di policy di Access Analyzer può aiutare a perfezionare le politiche di autorizzazione.
- Assicurati che gli utenti abbiano un accesso limitato agli ambienti di produzione: gli utenti devono avere accesso agli ambienti di produzione solo in presenza di un caso d'uso valido. Una volta eseguite le attività specifiche che richiedono l'accesso alla produzione, l'accesso dell'utente deve essere revocato. Limitare l'accesso agli ambienti di produzione contribuisce a evitare eventi indesiderati con impatto sulla produzione e contiene gli effetti di accessi involontari.
- Prendi in considerazione i limiti delle autorizzazioni: un limite di autorizzazioni è una funzionalità che consente di utilizzare una politica gestita che imposta le autorizzazioni massime che una politica basata sull'identità può concedere a un'entità. IAM Il limite delle autorizzazioni di un'entità consente di eseguire solo le operazioni consentite dalle sue policy basate su identità e dai suoi limiti delle autorizzazioni.
- Prendi in considerazione i tag delle risorse per le autorizzazioni: puoi concedere l'accesso in base allo scopo della risorsa, al proprietario, all'ambiente o ad altri criteri servendoti di un modello di controllo degli accessi basato sugli attributi che utilizza i tag delle risorse. Ad esempio, puoi usare tag di risorse per diversificare gli ambienti di produzione e sviluppo. Tramite questi tag puoi limitare

gli sviluppatori all'ambiente di sviluppo. Abbinando policy su tag e autorizzazioni, puoi ottenere l'accesso a risorse dettagliate senza dover definire policy personalizzate e complesse per ogni funzione professionale.

- Utilizza AWS Organizations le politiche di controllo del servizio per. Le policy di controllo dei servizi monitorano a livello centrale il numero massimo di autorizzazioni disponibili per gli account membri della tua organizzazione. È importante notare che le policy di controllo dei servizi consentono di limitare le autorizzazioni dell'utente root negli account membri. Prendi in considerazione anche l'utilizzo AWS Control Tower, che fornisce controlli gestiti prescrittivi che arricchiscono. AWS Organizations Puoi anche definire i tuoi controlli in Control Tower.
- Stabilisci una politica del ciclo di vita degli utenti per la tua organizzazione: le politiche sul ciclo di vita degli utenti definiscono le attività da eseguire quando gli utenti vengono integrati AWS, cambiano ruolo o ambito lavorativo o non hanno più bisogno di accedervi. AWS Le revisioni delle autorizzazioni vanno eseguite in ogni fase del ciclo di vita di un utente per verificare che siano sufficientemente restrittive e per evitare che si accumulino.
- Stabilisci una pianificazione regolare per rivedere le autorizzazioni e rimuovere eventuali autorizzazioni non necessarie: dovresti controllare regolarmente l'accesso degli utenti per verificare che gli utenti non dispongano di un accesso eccessivamente permissivo. AWS Confige IAM Access Analyzer possono aiutarti nel controllo delle autorizzazioni degli utenti.
- Stabilisci una matrice dei ruoli lavorativi: una matrice dei ruoli lavorativi visualizza i vari ruoli e i livelli di accesso richiesti all'interno del tuo ambiente. AWS Tramite una matrice dei ruoli professionali puoi definire e separare le autorizzazioni in base alle responsabilità degli utenti all'interno dell'organizzazione. Utilizza i gruppi anziché applicare le autorizzazioni direttamente a singoli utenti o ruoli.

Risorse

Documenti correlati:

- Grant least privilege
- Limiti delle autorizzazioni per le entità IAM
- Tecniche per scrivere politiche sui privilegi minimi IAM
- IAMAccess Analyzer semplifica l'implementazione delle autorizzazioni con privilegi minimi generando IAM policy basate sull'attività di accesso
- Delega la gestione delle autorizzazioni agli sviluppatori utilizzando i limiti delle autorizzazioni IAM
- Perfezionare le autorizzazioni utilizzando le informazioni dell'ultimo accesso

- IAMtipi di policy e quando utilizzarle
- Test delle IAM politiche con il simulatore IAM di politiche
- Guardrail in AWS Control Tower
- Architetture Zero Trust: una prospettiva AWS
- Come implementare il principio del privilegio minimo con CloudFormation StackSets
- Controllo degli accessi basato sugli attributi () ABAC
- Reducing policy scope by viewing user activity
- View role access
- Uso dei tag per organizzare il proprio ambiente e aumentare la responsabilità
- Strategie di applicazione di tag AWS
- Applicazione di tag alle risorse AWS

Video correlati:

- Next-generation permissions management
- Zero Trust: una prospettiva AWS

Esempi correlati:

- Lab: IAM autorizzazioni, limiti, delega della creazione di ruoli
- Lab: controllo degli accessi basato su IAM tag per EC2

SEC03-BP03 Stabilire un processo di accesso di emergenza

Crea un processo che consenta l'accesso di emergenza ai tuoi carichi di lavoro nell'improbabile eventualità che si verifichi un problema con il tuo gestore dell'identità digitale centralizzato.

Devi progettare processi per diverse modalità di guasto che potrebbero causare un evento di emergenza. Ad esempio, in circostanze normali, gli utenti della forza lavoro si federano al cloud utilizzando un provider di identità centralizzato (SEC02-BP04) per gestire i propri carichi di lavoro. Tuttavia, se il tuo gestore dell'identità digitale centralizzato riscontra un errore o la configurazione per la federazione nel cloud subisce modifiche, gli utenti della tua forza lavoro potrebbero non essere in grado di federarsi nel cloud. Un processo di accesso di emergenza consente agli amministratori autorizzati di accedere alle risorse cloud tramite mezzi alternativi (come una forma alternativa di federazione o l'accesso diretto degli utenti) per risolvere problemi relativi alla configurazione della

federazione o ai carichi di lavoro. Si ricorre al processo di accesso di emergenza fino al ripristino del normale meccanismo di federazione.

Risultato desiderato:

- Hai definito e documentato le modalità di guasto che costituiscono un'emergenza: considera le circostanze normali e i sistemi da cui dipendono gli utenti per gestire i loro carichi di lavoro. Prendi in considerazione quali guasti possono interessare ciascuna di queste dipendenze e causare una situazione di emergenza. Potresti trovare utili le domande e le best practice del pilastro dell'affidabilità per individuare le modalità di errore e progettare sistemi più resilienti al fine di ridurre al minimo la probabilità di guasti.
- Hai documentato i passaggi da seguire per confermare che un guasto costituisce un'emergenza. Ad esempio, puoi richiedere agli amministratori di identità di controllare lo stato dei gestori delle identità digitali primari e di standby e, se entrambi non sono disponibili, dichiarare un evento di emergenza per guasto del gestore dell'identità digitale.
- È stato definito un processo di accesso di emergenza specifico per ogni tipo di modalità di emergenza o di guasto. Essere specifici può ridurre la tentazione da parte degli utenti di abusare di un processo generale per tutti i tipi di emergenze. I processi di accesso di emergenza illustrano le circostanze in cui ciascun processo va o non va utilizzato e indicano processi alternativi applicabili.
- I tuoi processi sono ben documentati con istruzioni e playbook dettagliati, facili da mettere in pratica in modo rapido ed efficiente. Ricorda che un evento di emergenza può essere un momento stressante per i tuoi utenti, che potrebbero essere sotto pressione per motivi di tempo, quindi progetta il tuo processo in modo che sia il più semplice possibile.

Anti-pattern comuni:

- Non si dispone di procedure di accesso di emergenza ben documentate e collaudate. Gli utenti non sono preparati per un'emergenza e seguono processi improvvisati quando si verifica un evento di emergenza.
- I processi di accesso di emergenza dipendono dagli stessi sistemi (come un gestore dell'identità digitale centralizzato) dei normali meccanismi di accesso. Ciò significa che il guasto di un sistema di questo tipo può influire sui normali meccanismi di accesso e di emergenza e compromettere la capacità di ripristino dall'errore.
- I processi di accesso di emergenza vengono utilizzati in situazioni non di emergenza. Ad esempio, gli utenti utilizzano spesso in modo improprio i processi di accesso di emergenza poiché trovano più facile apportare modifiche direttamente piuttosto che inviarle tramite una pipeline.

- I processi di accesso di emergenza non generano log sufficienti per effettuare l'audit dei processi oppure i log non vengono monitorati per segnalare un potenziale uso improprio dei processi.

Vantaggi dell'adozione di questa best practice:

- Grazie a processi di accesso di emergenza ben documentati e collaudati, puoi ridurre il tempo impiegato dagli utenti per rispondere a un evento di emergenza e risolverlo. Ciò può comportare una riduzione dei tempi di inattività e una maggiore disponibilità dei servizi forniti ai clienti.
- È possibile tenere traccia di ogni richiesta di accesso di emergenza e rilevare e segnalare i casi di tentativi non autorizzati di uso improprio del processo per eventi non di emergenza.

Livello di rischio associato se questa best practice non fosse adottata: medio

Guida all'implementazione

Questa sezione fornisce indicazioni per la creazione di processi di accesso di emergenza per diverse modalità di errore relative ai carichi di lavoro distribuiti AWS, a partire da linee guida comuni che si applicano a tutte le modalità di errore e seguite da linee guida specifiche basate sul tipo di modalità di errore.

Linee guida comuni per tutte le modalità di errore

Nella progettazione di un processo di accesso di emergenza per una modalità di errore, tieni presente quanto segue:

- Documenta prerequisiti e presupposti del processo: quando il processo deve e non deve essere utilizzato. Aiuta a descrivere in dettaglio la modalità di errore e a documentare le ipotesi, come lo stato di altri sistemi correlati. Ad esempio, il processo per la Failure Mode 2 presuppone che il provider di identità sia disponibile, ma la configurazione attivata AWS sia stata modificata o sia scaduta.
- Precrea le risorse necessarie per il processo di accesso di emergenza (SEC10-BP05). Ad esempio, precrea l'accesso di emergenza Account AWS con IAM utenti e ruoli e i ruoli tra account in tutti gli account del carico di lavoroIAM. Ciò assicura che queste risorse siano pronte e disponibili quando si verifica un evento di emergenza. Creando prima le risorse, non si ha alcuna dipendenza dal piano di AWS controllo APIs (utilizzato per creare e modificare AWS le risorse) che potrebbe non essere disponibile in caso di emergenza. Inoltre, precreando IAM le risorse, non è necessario tenere conto dei potenziali ritardi dovuti all'eventuale coerenza.

- Includi i processi di accesso di emergenza come parte dei tuoi piani di gestione degli incidenti (SEC10-BP02). Documenta le modalità in cui si tiene traccia degli eventi di emergenza e come questi vengono comunicati ad altri membri dell'organizzazione, come i team di pari livello, la leadership e, se applicabile, esternamente ai clienti e ai partner aziendali.
- Definisci il processo di richiesta di accesso di emergenza nel tuo sistema di flusso di lavoro esistente, se ne hai uno, per le richieste di assistenza. In genere, tali sistemi di flusso di lavoro consentono di creare moduli di acquisizione per raccogliere informazioni sulla richiesta, tenere traccia della richiesta in ogni fase del flusso di lavoro e aggiungere passaggi di approvazione automatici e manuali. Collega ciascuna richiesta a un evento di emergenza corrispondente tracciato nel tuo sistema di gestione degli incidenti. Disporre di un sistema uniforme per gli accessi di emergenza consente di tenere traccia di tali richieste in un unico sistema, analizzare le tendenze di utilizzo e migliorare i processi.
- Verifica che i processi di accesso di emergenza possano essere avviati solo da utenti autorizzati e richiedano l'approvazione di colleghi o manager dell'utente, a seconda dei casi. Il processo di approvazione deve funzionare in modo efficacie sia all'interno sia al di fuori dell'orario lavorativo. Definisci in che modo le richieste di approvazione possono essere eseguite da approvatori secondari, qualora gli approvatori principali non fossero disponibili, e come vengono inoltrate lungo la catena di gestione fino all'approvazione.
- Verifica che il processo generi log di audit ed eventi dettagliati per i tentativi riusciti e non andati a buon fine di ottenere l'accesso di emergenza. Monitora sia il processo di richiesta sia il meccanismo di accesso di emergenza per rilevare usi impropri o accessi non autorizzati. Metti in correlazione l'attività con gli eventi di emergenza in corso dal tuo sistema di gestione degli incidenti e segnala i casi in cui le azioni si verificano al di fuori dei periodi di tempo previsti. Ad esempio, devi monitorare e inviare avvisi in merito ad attività nell' Account AWS di accesso di emergenza, poiché non dovrebbe mai essere utilizzato per le normali operazioni.
- Testa periodicamente i processi di accesso di emergenza per verificare che i passaggi siano chiari e garantire il livello di accesso corretto in modo rapido ed efficiente. I processi di accesso di emergenza devono essere testati nell'ambito delle simulazioni di risposta agli incidenti (SEC10-BP07) e dei test di disaster recovery (-BP03). REL13

Modalità di errore 1: il provider di identità utilizzato per la federazione non è disponibile AWS

Come descritto in SEC02-BP04 Affidati a un provider di identità centralizzato, ti consigliamo di affidarti a un provider di identità centralizzato per federare la forza lavoro a cui concedere l'accesso agli utenti. Account AWSÈ possibile eseguire la federazione tra più membri AWS dell'organizzazione utilizzando IAM Identity Center Account AWS oppure è possibile eseguire la federazione per uso

individuale. Account AWS IAM In entrambi i casi, gli utenti della forza lavoro si autenticano con il gestore dell'identità digitale centralizzato prima di essere reindirizzati a un endpoint di accesso AWS per l'autenticazione unica.

Nell'improbabile eventualità che il gestore dell'identità digitale centralizzato non sia disponibile, gli utenti della tua forza lavoro non possono federarsi per accedere agli Account AWS o gestire i propri carichi di lavoro. In questo caso di emergenza, puoi fornire una procedura di accesso di emergenza a cui un piccolo gruppo di amministratori può accedere per Account AWS eseguire attività critiche che non possono attendere che i provider di identità centralizzati tornino online. Ad esempio, il tuo provider di identità non è disponibile per 4 ore e durante quel periodo devi modificare i limiti superiori di un gruppo Amazon EC2 Auto Scaling in un account di produzione per gestire un picco imprevisto nel traffico dei clienti. Gli amministratori di emergenza devono seguire la procedura di accesso di emergenza per accedere alla produzione specifica Account AWS e apportare le modifiche necessarie.

Il processo di accesso di emergenza si basa su un accesso di emergenza precreato Account AWS che viene utilizzato esclusivamente per l'accesso di emergenza e dispone di AWS risorse (come IAM ruoli e IAM utenti) per supportare il processo di accesso di emergenza. Durante le normali operazioni, nessuno deve accedere all'account di accesso di emergenza ed è necessario monitorare e fornire avvisi riguardo a usi impropri di questo account (per maggiori dettagli, vedi la sezione precedente Linee guida comuni).

L'account per l'accesso di emergenza dispone di IAM ruoli di accesso di emergenza con autorizzazioni per assumere ruoli tra account diversi in quelli Account AWS che richiedono l'accesso di emergenza. Questi IAM ruoli sono precreati e configurati con politiche di fiducia che considerano attendibili i ruoli dell'account di emergenza. IAM

Per il processo di accesso di emergenza è possibile utilizzare uno dei seguenti approcci:

- È possibile precreare un set di IAMutenti per gli amministratori di emergenza nell'account di accesso di emergenza con password e token complessi associati. MFA Questi IAM utenti dispongono delle autorizzazioni per assumere i IAM ruoli che consentono quindi l'accesso da più account all'area in cui è richiesto l' Account AWS accesso di emergenza. Ti consigliamo di creare il minor numero possibile di utenti di questo tipo e di assegnare ogni utente a un unico amministratore di emergenza. Durante un'emergenza, un utente amministratore di emergenza accede all'account di accesso di emergenza utilizzando la propria password e il codice MFA token, passa al IAM ruolo di accesso di emergenza nell'account di emergenza e infine passa al IAM ruolo di accesso di emergenza nell'account del carico di lavoro per eseguire l'azione di modifica di emergenza. Il vantaggio di questo approccio è che ogni IAM utente viene assegnato

a un amministratore di emergenza ed è possibile sapere quale utente ha effettuato l'accesso esaminando gli eventi. CloudTrail Lo svantaggio è che è necessario mantenere più IAM utenti con le password e i token di lunga durata associati. MFA

- È possibile utilizzare l'utente Account AWS root per l'accesso di emergenza per accedere all'account di accesso di emergenza, assumere il IAM ruolo di accesso di emergenza e assumere il ruolo multiaccount nell'account del carico di lavoro. Ti consigliamo di impostare una password sicura e più MFA token per l'utente root. Consigliamo inoltre di archiviare la password e i MFA token in un archivio di credenziali aziendali sicuro che applichi un'autenticazione e un'autorizzazione avanzate. È necessario proteggere i fattori di reimpostazione della password e del MFA token: imposta l'indirizzo e-mail dell'account in una lista di distribuzione e-mail monitorata dagli amministratori della sicurezza del cloud e il numero di telefono dell'account su un numero di telefono condiviso monitorato anche dagli amministratori della sicurezza. Il vantaggio di questo approccio è l'esistenza di un solo set di credenziali utente root da gestire. Lo svantaggio è che, trattandosi di un utente condiviso, più amministratori hanno la possibilità di accedere come utente root. Controlla gli eventi del log del tuo vault aziendale per identificare quale amministratore ha utilizzato la password dell'utente root.

Modalità di errore 2: la configurazione del provider di identità attiva AWS è stata modificata o è scaduta

Per consentire agli utenti della forza lavoro di effettuare la federazione Account AWS, è possibile configurare l'IAMIdentity Center con un provider di identità esterno o creare un IAM provider di identità (SEC02-BP04). In genere, li configuri importando un documento di SAML XML metadati fornito dal tuo provider di identità. Il XML documento di metadati include un certificato X.509 corrispondente a una chiave privata utilizzata dal provider di identità per firmare le proprie asserzioni. SAML

Queste configurazioni sul AWS lato -possono essere modificate o eliminate per errore da un amministratore. In un altro scenario, il certificato X.509 importato in AWS potrebbe scadere e i nuovi metadati XML con un nuovo certificato non sono ancora stati importati in. AWS Entrambi gli scenari possono interrompere la federazione degli utenti della forza lavoro, con conseguente emergenza. AWS

In un caso di emergenza di questo tipo, puoi fornire agli amministratori delle identità l'accesso AWS a cui risolvere i problemi di federazione. Ad esempio, l'amministratore delle identità utilizza la procedura di accesso di emergenza per accedere all'accesso di emergenza Account AWS, passa a un ruolo nell'account amministratore dell'Identity Center e aggiorna la configurazione del provider di identità

esterno importando il XML documento di SAML metadati più recente dal provider di identità per riattivare la federazione. Una volta ristabilita la federazione, gli utenti della forza lavoro continuano a utilizzare il normale processo operativo per federare l'accesso ai propri account di carico di lavoro.

È possibile seguire gli approcci illustrati nella sezione precedente Modalità di errore 1 per creare un processo di accesso di emergenza. Puoi concedere le autorizzazioni con il privilegio minimo agli amministratori delle identità per accedere solo all'account amministratore di Centro identità ed eseguire azioni in Centro identità in quell'account.

Modalità di errore 3: blocco del Centro identità

Nell'improbabile eventualità di un IAM Identity Center o di un' Regione AWS interruzione, ti consigliamo di configurare una configurazione da utilizzare per fornire un accesso temporaneo a. AWS Management Console

Il processo di accesso di emergenza utilizza la federazione diretta dal provider di identità IAM a un account di emergenza. Per informazioni dettagliate sul processo e sulle considerazioni di progettazione, consulta Set up emergency access to the AWS Management Console.

Passaggi dell'implementazione

Passaggi comuni per tutte le modalità di errore

- Creane uno Account AWS dedicato ai processi di accesso di emergenza. Precrea le IAM risorse necessarie nell'account, ad esempio IAM ruoli o IAM utenti e, facoltativamente, gli IAM Identity Provider. Inoltre, crea in anticipo IAM ruoli interaccount nel carico di lavoro con relazioni di fiducia Account AWS con i IAM ruoli corrispondenti nell'account per l'accesso di emergenza. Puoi utilizzare AWS CloudFormation StackSets with AWS Organizations per creare tali risorse negli account dei membri della tua organizzazione.

- Crea politiche di controllo del AWS Organizations servizio (SCPs) per negare l'eliminazione e la modifica dei IAM ruoli tra account diversi nel membro. Account AWS
- Abilita CloudTrail l'accesso di emergenza Account AWS e invia gli eventi del percorso a un bucket S3 centrale nella tua raccolta di log. Account AWS Se lo utilizzi AWS Control Tower per configurare e gestire il tuo ambiente AWS multi-account, ogni account che crei utilizzando AWS Control Tower o a cui ti registri AWS Control Tower è CloudTrail abilitato per impostazione predefinita e inviato a un bucket S3 in un archivio di log dedicato. Account AWS
- Monitora l'attività dell'account con accesso di emergenza creando EventBridge regole che corrispondano all'accesso alla console e all'APIattività dei ruoli di emergenza. IAM Invia notifiche al

tuo centro operativo di sicurezza quando si verificano attività al di fuori di un evento di emergenza in corso e di cui hai traccia nel tuo sistema di gestione degli incidenti.

Passaggi aggiuntivi per la modalità di errore 1: il provider di identità utilizzato per la federazione non AWS è disponibile e la modalità di errore 2: la configurazione del provider di identità attiva AWS è stata modificata o è scaduta

- Crea preliminarmente le risorse in base al meccanismo scelto per l'accesso di emergenza:
  - Utilizzo IAM degli utenti: precrea gli IAM utenti con password complesse e dispositivi associati. MFA
  - Utilizzando l'utente utente root dell'account di emergenza: configura l'utente root con una password sicura e archivia la password nel tuo vault di credenziali aziendali. Associa più MFA dispositivi fisici all'utente root e archivia i dispositivi in posizioni a cui possono accedere rapidamente i membri del team di amministrazione delle emergenze.

Passaggi aggiuntivi per la Modalità di errore 3: blocco del Centro identità

- Come descritto in dettaglio in Configurazione dell'accesso di emergenza AWS Management Console, in caso di accesso di emergenza Account AWS, crea un provider di IAM identità per consentire la SAML federazione diretta dal tuo provider di identità.
- Crea gruppi operativi di emergenza nel tuo IdP senza membri.
- Crea IAM ruoli corrispondenti ai gruppi operativi di emergenza nell'account per l'accesso di emergenza.

Risorse

Best practice Well-Architected correlate:

- SEC02-BP04 Affidati a un provider di identità centralizzato
- SEC03-BP02 Concedi l'accesso con privilegi minimi
- SEC10-BP02 Sviluppare piani di gestione degli incidenti
- SEC10-BP07 Run game days

Documenti correlati:

- Configurare l'accesso di emergenza al AWS Management Console

- Consentire agli utenti federati SAML 2.0 di accedere a AWS Management Console
- Break glass access

Video correlati:

- AWS re:Invent 2022 - Semplifica l'accesso alla forza lavoro esistente con Identity Center IAM
- AWS re:Inforce 2022 - () approfondimento AWS Identity and Access Management IAM

Esempi correlati:

- AWS Break Glass Role
- Framework AWS per playbook per i clienti
- AWS incident response playbook samples

SEC03-BP04 Ridurre continuamente le autorizzazioni

Man mano che i team determinano gli accessi necessari, rimuovi le autorizzazioni non necessarie e stabilisci processi di revisione per ottenere le autorizzazioni con il privilegio minimo. Monitora costantemente e rimuovi le identità e le autorizzazioni inutilizzate per l'accesso sia umano che delle macchine.

Risultato desiderato: le policy di autorizzazione rispettano il principio del privilegio minimo. Man mano che le mansioni e i ruoli vengono definiti meglio, è necessario rivedere le policy di autorizzazione per eliminare le autorizzazioni non necessarie. Questo approccio riduce la portata dell'impatto nel caso di esposizione accidentale delle credenziali o di accesso in altro modo senza autorizzazione.

Anti-pattern comuni:

- L'impostazione predefinita è la concessione delle autorizzazioni di amministratore agli utenti.
- Creazione di policy eccessivamente permissive, ma senza privilegi completi di amministratore.
- Mantenimento delle policy di autorizzazione anche quando non sono più necessarie.

Livello di rischio associato se questa best practice non fosse adottata: medio

Guida all'implementazione

Quando i team e i progetti sono in fase iniziale, è possibile usare policy di autorizzazione permissiva per stimolare l'innovazione e l'agilità. Ad esempio, in un ambiente di sviluppo o di test, gli sviluppatori possono avere accesso a un'ampia gamma di servizi. AWS Si consiglia di valutare in modo costante gli accessi e di limitare l'accesso solo ai servizi e alle azioni di servizio necessari per completare il lavoro in corso. Raccomandiamo questa valutazione sia per l'identità umana che per quella macchina. Le identità delle macchine, a volte denominate account di sistema o di servizio, sono identità che consentono AWS l'accesso ad applicazioni o server. Questo accesso è particolarmente importante in un ambiente di produzione, dove autorizzazioni troppo permissive possono avere un ampio impatto e potenzialmente esporre i dati dei clienti.

AWS fornisce diversi metodi per aiutare a identificare utenti, ruoli, autorizzazioni e credenziali non utilizzati. AWS può anche aiutare ad analizzare l'attività di accesso di IAM utenti e ruoli, incluse le chiavi di accesso associate, e l'accesso a AWS risorse come oggetti nei bucket Amazon S3. AWS Identity and Access Management Access Analyzer la generazione di policy può aiutarti a creare policy di autorizzazione restrittive basate sui servizi e sulle azioni effettivi con cui interagisce un preside. Il controllo degli accessi basato sugli attributi (ABAC) può contribuire a semplificare la gestione delle autorizzazioni, in quanto è possibile fornire le autorizzazioni agli utenti utilizzando i relativi attributi anziché allegare le politiche di autorizzazione direttamente a ciascun utente.

Passaggi dell'implementazione

- Utilizzo AWS Identity and Access Management Access Analyzerhttps://docs.aws.amazon.com/IAM/latest/UserGuide/what-is-access-analyzer.html: IAM Access Analyzer aiuta a identificare le risorse dell'organizzazione e degli account, come i bucket o i IAM ruoli di Amazon Simple Storage Service (Amazon S3) condivisi con un'entità esterna.
- Usa la generazione di policy di IAM Access Analyzer: la generazione di policy di IAM Access Analyzer ti aiuta a creare policy di autorizzazione granulari basate sull'attività di accesso di un utente o di un ruolo. IAM
- Determina un periodo di tempo e una politica di utilizzo accettabili per IAM utenti e ruoli: utilizza il timestamp dell'ultimo accesso per identificare gli utenti e i ruoli non utilizzati e rimuoverli. Rivedi le informazioni sull'ultimo accesso al servizio e sull'ultima azione per identificare e definire le autorizzazioni per specifici utenti e ruoli. Ad esempio, puoi utilizzare le informazioni sull'ultimo accesso per identificare le azioni specifiche di Amazon S3 richieste dal ruolo dell'applicazione e delimitare l'accesso del ruolo solo a tali azioni. Le funzionalità relative alle informazioni relative all'ultimo accesso sono disponibili in AWS Management Console e consentono di incorporarle a livello di codice nei flussi di lavoro e negli strumenti automatizzati dell'infrastruttura.

- Prendi in considerazione la possibilità di registrare gli eventi relativi ai dati in AWS CloudTrail: Per impostazione predefinita, CloudTrail non registra eventi di dati come le attività a livello di oggetto di Amazon S3 (ad esempio`GetObject`, and) `DeleteObject` o le attività di tabella Amazon DynamoDB (ad esempio e). `PutItem` `DeleteItem` Considera l'uso della creazione di log di questi eventi per stabilire quali utenti e ruoli devono accedere a specifici oggetti Amazon S3 o elementi di tabelle DynamoDB.

Risorse

Documenti correlati:

- Grant least privilege
- Rimuovere credenziali non necessarie
- Che cos'è? AWS CloudTrail
- Lavorare con le policy
- Logging and monitoring DynamoDB
- Utilizzo della registrazione CloudTrail degli eventi per bucket e oggetti Amazon S3
- Ottenere report sulle credenziali per il tuo Account AWS

Video correlati:

- Diventa un IAM Policy Master in 60 minuti o meno
- Separation of Duties, Least Privilege, Delegation, and CI/CD
- AWS Re:inforce 2022 - AWS Identity and Access Management () IAM approfondimento

SEC03-BP05 Definisci barriere di autorizzazione per la tua organizzazione

Utilizza i guardrail delle autorizzazioni per ridurre l'ambito delle autorizzazioni disponibili concedibili ai principali. La catena di valutazione delle policy di autorizzazione comprende i guardrail così da determinare le autorizzazioni effettive di un principale quando adotta decisioni relative alle autorizzazioni. È possibile definire i guardrail utilizzando un approccio basato sui livelli. Applica alcuni guardrail in modo esteso all'intera organizzazione e applicane altri in modo granulare alle sessioni di accesso temporaneo.

Risultato desiderato: hai un chiaro isolamento degli ambienti utilizzando Account AWS separati. Le politiche di controllo del servizio (SCPs) vengono utilizzate per definire barriere di autorizzazione a

livello di organizzazione. I guardrail più estesi sono impostati ai livelli gerarchici più vicini alla radice dell'organizzazione, mentre i guardrail più rigidi sono impostati più vicino al livello dei singoli account. Se supportate, le policy sulle risorse definiscono le condizioni che un principale deve soddisfare per ottenere l'accesso a una risorsa. Le policy per le risorse, inoltre, definiscono l'insieme delle azioni consentite, laddove appropriato. I limiti delle autorizzazioni sono posti sui principali che gestiscono le autorizzazioni del carico di lavoro, delegando la gestione delle autorizzazioni ai singoli proprietari del carico di lavoro.

Anti-pattern comuni:

- Creazione di membri Account AWS all'interno di un'AWS organizzazione, ma non utilizzo SCPs per limitare l'uso e le autorizzazioni disponibili alle relative credenziali root.
- Assegnare le autorizzazioni in base al privilegio minimo, senza però porre guardrail sull'insieme massimo di autorizzazioni concedibili.
- Affidarsi alla base di negazione implicita AWS IAM per limitare le autorizzazioni, confidando che le politiche non concedano autorizzazioni di autorizzazione esplicite indesiderate.
- Eseguiamo più ambienti di carico di lavoro nello stesso ambiente e poi Account AWS facciamo affidamento su meccanismi come tag o politiche delle risorse per far rispettare i limiti delle VPCs autorizzazioni.

Vantaggi derivanti dall'adozione di questa best practice: i guardrail di autorizzazione contribuiscono a creare la certezza che le autorizzazioni indesiderate non possano essere concesse, anche quando una policy di autorizzazione tenta di farlo. Ciò può semplificare la definizione e la gestione delle autorizzazioni riducendo l'ambito massimo delle autorizzazioni da prendere in considerazione.

Livello di rischio associato se questa best practice non fosse adottata: medio

Guida all'implementazione

Ti consigliamo di utilizzare un approccio basato sui livelli per definire i guardrail di autorizzazione per la tua organizzazione. Questo approccio riduce in modo sistematico il set massimo di autorizzazioni possibili con l'applicazione di livelli aggiuntivi. Ciò consente di concedere l'accesso in base al principio del privilegio minimo, riducendo il rischio di accessi non intenzionali dovuti a un'errata configurazione delle policy.

Il primo passo per definire i guardrail delle autorizzazioni è isolare i carichi di lavoro e gli ambienti in Account AWS separati. I responsabili di un account non possono accedere alle risorse di un altro

account senza l'autorizzazione esplicita in tal senso, anche se entrambi gli account fanno parte della stessa AWS organizzazione o fanno parte della stessa unità organizzativa (OU). Puoi utilizzarli OUs per raggruppare gli account che desideri amministrare come una singola unità.

Il passaggio successivo consiste nel ridurre il set massimo di autorizzazioni che è possibile concedere ai principali all'interno degli account dei membri dell'organizzazione. A tale scopo è possibile utilizzare le politiche di controllo del servizio (SCPs), che è possibile applicare a un'unità organizzativa o a un account. SCPspuò applicare controlli di accesso comuni, ad esempio limitare l'accesso a determinati elementi Regioni AWS, impedire l'eliminazione di risorse o disabilitare azioni di servizio potenzialmente rischiose. SCPsle informazioni applicate alla radice dell'organizzazione influiscono solo sugli account dei membri, non sull'account di gestione. SCPsgestisci solo i dirigenti all'interno della tua organizzazione. SCPsNon siete i dirigenti esterni all'organizzazione che accedono alle vostre risorse.

Un ulteriore passo consiste nell'utilizzare le politiche in materia di IAM risorse per definire le azioni disponibili che è possibile intraprendere sulle risorse da essi governate, oltre alle condizioni che il responsabile ad interim deve soddisfare. Ciò può essere tanto ampio quanto consentire tutte le azioni purché il responsabile faccia parte dell'organizzazione (utilizzando la chiave di PrincipalOrgId condizione), oppure granulare, consentire solo azioni specifiche per un ruolo specificoIAM. È possibile adottare un approccio simile con le condizioni nelle politiche di fiducia dei IAM ruoli. Se una politica di fiducia in materia di risorse o ruoli nomina esplicitamente un responsabile nello stesso account del ruolo o della risorsa da essa governata, tale responsabile non necessita di una IAM policy allegata che conceda le stesse autorizzazioni. Se il responsabile si trova in un account diverso da quello della risorsa, allora ha bisogno di una IAM politica allegata che conceda tali autorizzazioni.

Spesso, un team addetto al carico di lavoro vorrà gestire le autorizzazioni richieste dal proprio carico di lavoro. Ciò potrebbe richiedere la creazione di nuovi IAM ruoli e politiche di autorizzazione. È possibile acquisire l'ambito massimo di autorizzazioni che il team può concedere in un limite di IAM autorizzazione e associare questo documento a un IAM ruolo che il team può quindi utilizzare per gestire i propri IAM ruoli e le proprie autorizzazioni. Questo approccio può fornire loro la possibilità di completare il proprio lavoro riducendo al contempo i rischi legati all'accesso amministrativo. IAM

Un passaggio più granulare consiste nell'implementazione di tecniche di gestione degli accessi privilegiati (PAM) e di gestione temporanea degli accessi elevati (). TEAM Un esempio PAM è quello di richiedere ai responsabili di eseguire l'autenticazione a più fattori prima di intraprendere azioni privilegiate. Per ulteriori informazioni, vedere Configurazione MFA dell'accesso protetto. API TEAMrichiede una soluzione che gestisca l'approvazione e il periodo di tempo entro il quale un principale può avere un accesso elevato. Un approccio consiste nell'aggiungere temporaneamente

il responsabile alla politica di fiducia del ruolo per un IAM ruolo con accesso elevato. Un altro approccio consiste nel ridurre, in condizioni normali, le autorizzazioni concesse a un responsabile da un IAM ruolo utilizzando una politica di sessione e quindi revocare temporaneamente questa restrizione durante la finestra temporale approvata. Per ulteriori informazioni sulle soluzioni convalidate da AWS e da alcuni partner selezionati, consulta Temporary elevated access.

Passaggi dell'implementazione

1. Isola i carichi di lavoro e gli ambienti in Account AWS separati.
2. SCPsUtilizzatelo per ridurre il set massimo di autorizzazioni che possono essere concesse ai responsabili all'interno degli account membri dell'organizzazione.
   a. Ti consigliamo di adottare un approccio di tipo allowlist nella stesura del documento, SCPs che neghi tutte le azioni tranne quelle consentite e le condizioni in base alle quali sono consentite. Inizia definendo le risorse che desideri controllare e imposta l'effetto su Deny. Utilizzate l' NotAction elemento per negare tutte le azioni tranne quelle specificate. Combinalo con una NotLike Condizione per definire quando sono consentite queste azioni, se applicabili, come StringNotLike e ArnNotLike.
   b. Consulta Service control policy examples.
3. Utilizza le politiche relative alle IAM risorse per definire e specificare le condizioni per le azioni consentite sulle risorse. Utilizza le condizioni nelle politiche di fiducia dei IAM ruoli per creare restrizioni all'assunzione dei ruoli.
4. Assegna limiti di IAM autorizzazione ai IAM ruoli che i team addetti al carico di lavoro possono quindi utilizzare per gestire i ruoli e le autorizzazioni dei propri carichi di lavoroIAM.
5. Valuta le TEAM soluzioni PAM in base alle tue esigenze.

Risorse

Documenti correlati:

- Perimetri di dati su AWS
- Establish permissions guardrails using data perimeters
- Logica di valutazione delle policy

Esempi correlati:

- Service control policy examples

Strumenti correlati:

- AWS Solution: Temporary Elevated Access Management
- Soluzioni partner di sicurezza convalidate per TEAM

SEC03-BP06 Gestisci l'accesso in base al ciclo di vita

Monitora e regola le autorizzazioni concesse ai tuoi principali (utenti, ruoli e gruppi) durante il loro ciclo di vita all'interno dell'organizzazione. Adatta le appartenenze ai gruppi quando gli utenti cambiano ruolo e rimuovi l'accesso quando un utente lascia l'organizzazione.

Risultato desiderato: monitori e modifichi le autorizzazioni durante l'intero ciclo di vita dei principali all'interno dell'organizzazione, riducendo così il rischio di privilegi superflui. Concedi l'accesso appropriato quando crei un utente. L'accesso viene modificato man mano che cambiano le responsabilità dell'utente e lo si rimuove quando l'utente non è più attivo o ha lasciato l'organizzazione. Gestisci a livello centrale le modifiche ai tuoi utenti, ruoli e gruppi. Utilizzate l'automazione per propagare le modifiche ai vostri ambienti. AWS

Anti-pattern comuni:

- Concedi alle identità privilegi di accesso eccessivi o estesi, al di là di quanto richiesto inizialmente.
- I privilegi di accesso non vengono rivisti e modificati poiché i ruoli e le responsabilità delle identità cambiano nel tempo.
- Le identità inattive o terminate vengono lasciate con privilegi di accesso attivi. Ciò aumenta il rischio di accessi non autorizzati.
- La gestione del ciclo di vita dell'identità non viene automatizzata.

Livello di rischio associato se questa best practice non fosse adottata: medio

Guida all'implementazione

Gestisci e adatta attentamente i privilegi di accesso che concedi alle identità (come utenti, ruoli, gruppi) durante il loro ciclo di vita. Questo ciclo di vita include la fase iniziale di onboarding, i continui cambiamenti di ruoli e responsabilità e l'eventuale offboarding o cessazione. Gestisci in modo proattivo l'accesso in base alla fase del ciclo di vita per mantenere il livello di accesso appropriato. Rispetta il principio del privilegio minimo per ridurre il rischio di privilegi di accesso eccessivi o non necessari.

È possibile gestire il ciclo di vita degli IAM utenti direttamente all'interno o tramite la federazione Account AWS, dal provider di identità della forza lavoro a Identity Center. AWS IAM Per IAM gli utenti, è possibile creare, modificare ed eliminare utenti e le relative autorizzazioni associate all'interno di. Account AWS Per gli utenti federati, è possibile utilizzare IAM Identity Center per gestirne il ciclo di vita sincronizzando le informazioni su utenti e gruppi dal provider di identità dell'organizzazione utilizzando il protocollo System for Cross-domain Identity Management (). SCIM

SCIMè un protocollo standard aperto per il provisioning e il disprovisioning automatici delle identità degli utenti su diversi sistemi. Integrando il proprio provider di IAM identità con Identity Center usingSCIM, è possibile sincronizzare automaticamente le informazioni su utenti e gruppi, contribuendo a verificare che i privilegi di accesso vengano concessi, modificati o revocati in base alle modifiche nella fonte di identità autorevole dell'organizzazione.

Man mano che i ruoli e le responsabilità dei dipendenti cambiano all'interno dell'organizzazione, modifica di conseguenza i loro privilegi di accesso. È possibile utilizzare i set di autorizzazioni di IAM Identity Center per definire diversi ruoli o responsabilità professionali e associarli alle politiche e alle autorizzazioni appropriate. IAM Quando il ruolo di un dipendente cambia, puoi aggiornare il set di autorizzazioni assegnato per riflettere le nuove responsabilità. Verifica che il dipendente disponga dell'accesso necessario rispettando il principio del privilegio minimo.

Passaggi dell'implementazione

1. Definisci e documenta un processo del ciclo di vita della gestione degli accessi, comprese le procedure per la concessione dell'accesso iniziale, le revisioni periodiche e l'offboarding.
2. Implementa IAM ruoli, gruppi e limiti di autorizzazioni per gestire l'accesso collettivamente e applicare i livelli di accesso massimi consentiti.
3. Esegui l'integrazione con un provider di identità federato (come Microsoft Active Directory, Okta, Ping Identity) come fonte autorevole per le informazioni su utenti e gruppi utilizzando Identity Center. IAM
4. Utilizza il SCIM protocollo per sincronizzare le informazioni su utenti e gruppi dal provider di identità nell'Identity Store di IAM Identity Center.
5. Crea set di autorizzazioni in IAM Identity Center che rappresentano diversi ruoli o responsabilità professionali all'interno dell'organizzazione. Definisci le IAM politiche e le autorizzazioni appropriate per ogni set di autorizzazioni.
6. Implementa revisioni regolari degli accessi, la relativa revoca tempestiva e il miglioramento continuo del processo del ciclo di vita della gestione degli accessi.
7. Offri formazione e sensibilizza i dipendenti in materia di best practice sulla gestione degli accessi.

Risorse

Best practice correlate:

- SEC02-BP04 Affidati a un provider di identità centralizzato

Documenti correlati:

- Manage your identity source
- Gestisci le identità in Identity Center IAM
- Uso di AWS Identity and Access Management Access Analyzer
- IAMGenerazione di policy di Access Analyzer

Video correlati:

- AWS RE:Inforce 2023 - Gestisci l'accesso temporaneo elevato con Identity Center AWS IAM
- AWS re:Invent 2022 - Semplifica l'accesso alla forza lavoro esistente con Identity Center IAM
- AWS re:Invent 2022 - Sfrutta la potenza delle politiche e limita le autorizzazioni con Access Analyzer IAM

SEC03-BP07 Analizza l'accesso pubblico e tra account

Monitora continuamente i risultati che evidenziano l'accesso multi-account e pubblico. Limita l'accesso multi-account e pubblico alle risorse che lo richiedono.

Risultato desiderato: scopri quali delle tue AWS risorse sono condivise e con chi. Monitora e sottoponi costantemente ad audit le risorse condivise per verificare che siano condivise solo con i principali autorizzati.

Anti-pattern comuni:

- Assenza di un inventario delle risorse condivise.
- Mancanza di un processo di approvazione dell'accesso multi-account e dell'accesso pubblico alle risorse.

Livello di rischio associato se questa best practice non fosse adottata: basso

Guida all'implementazione

Se il tuo account è attivo AWS Organizations, puoi concedere l'accesso alle risorse all'intera organizzazione, a unità organizzative specifiche o a singoli account. Se l'account non è membro di un'organizzazione, puoi condividere le risorse con account individuali. Puoi concedere l'accesso diretto a più account utilizzando politiche basate sulle risorse, ad esempio le policy bucket di Amazon Simple Storage Service (Amazon S3), o consentendo a un responsabile di un altro account di assumere un ruolo nel tuo account. IAM Quando utilizzi le policy sulle risorse, verifica che l'accesso sia concesso solo ai principali autorizzati. Definisci un processo per approvare tutte le risorse che devono essere pubblicamente disponibili.

AWS Identity and Access Management Access Analyzer utilizza una sicurezza comprovabile per identificare tutti i percorsi di accesso a una risorsa dall'esterno del proprio account. Esamina continuamente le policy delle risorse e segnala i risultati dell'accesso multi-account e pubblico per semplificare l'analisi di accessi potenzialmente estesi. Prendi in considerazione la possibilità di configurare IAM Access Analyzer con AWS Organizations per verificare di avere visibilità su tutti i tuoi account. IAMAccess Analyzer consente inoltre di visualizzare in anteprima i risultati prima di distribuire le autorizzazioni per le risorse. In questo modo è possibile verificare che le modifiche alle policy garantiscano alle risorse solo l'accesso multi-account e pubblico previsto. In caso di progettazione per l'accesso multi-account, puoi utilizzare policy di affidabilità per controllare i casi in cui è possibile assumere un ruolo. Ad esempio, puoi utilizzare la chiave di condizione `PrincipalOrgId` per negare un tentativo di assumere un ruolo al di fuori di AWS Organizations.

AWS Config è in grado di segnalare le risorse non configurate correttamente e, tramite controlli delle AWS Config politiche, è in grado di rilevare le risorse per le quali è configurato l'accesso pubblico. Servizi come AWS Control Towere AWS Security Hubsemplificano l'implementazione di controlli investigativi e barriere AWS Organizations per identificare e riparare le risorse esposte al pubblico. Ad esempio, AWS Control Tower dispone di un guardrail gestito in grado di rilevare se alcune EBSistantanee di Amazon sono ripristinabili da. Account AWS

Passaggi dell'implementazione

- Prendi in considerazione l'utilizzo di AWS Config for AWS Organizations: AWS Config consente di aggregare i risultati di più account all'interno di un account AWS Organizations amministratore delegato. Ciò fornisce una visione completa e consente di eseguire la distribuzione Regole di AWS Config su più account per rilevare risorse accessibili al pubblico.

- Configure AWS Identity and Access Management Access Analyzer IAM Access Analyzer ti aiuta a identificare le risorse della tua organizzazione e dei tuoi account, come i bucket Amazon S3 IAM o i ruoli condivisi con un'entità esterna.
- Usa la riparazione automatica AWS Config per rispondere alle modifiche nella configurazione dell'accesso pubblico dei bucket Amazon S3: puoi attivare automaticamente le impostazioni di accesso pubblico a blocchi per i bucket Amazon S3.
- Implementa monitoraggio e avvisi per stabilire se i bucket Amazon S3 sono diventati pubblici: devi disporre di monitoraggio e avvisi per stabilire se il blocco dell'accesso pubblico Amazon S3 è disattivato e se i bucket Amazon S3 diventano pubblici. Inoltre, se lo utilizzi AWS Organizations, puoi creare una policy di controllo del servizio che impedisca modifiche alle policy di accesso pubblico di Amazon S3. AWS Trusted Advisor verifica la presenza di bucket Amazon S3 con autorizzazioni di accesso aperto. Le autorizzazioni bucket che concedono, caricano o eliminano l'accesso per chiunque danno origine a potenziali problemi di sicurezza, consentendo a chiunque di aggiungere, modificare o rimuovere elementi in un bucket. Il Trusted Advisor controllo esamina le autorizzazioni esplicite dei bucket e le politiche associate ai bucket che potrebbero avere la precedenza sulle autorizzazioni dei bucket. Puoi anche utilizzarli AWS Config per monitorare i bucket Amazon S3 per l'accesso pubblico. Per ulteriori informazioni, consulta Come utilizzare AWS Config per monitorare e rispondere ai bucket Amazon S3 che consentono l'accesso pubblico. Durante la revisione dell'accesso, è importante prendere in considerazione i tipi di dati contenuti nei bucket Amazon S3. Amazon Macie aiuta a scoprire e proteggere dati sensibili, ad esempio, e credenzialiPHI, come dati privati o chiavi. PII AWS

Risorse

Documenti correlati:

- Uso di AWS Identity and Access Management Access Analyzer
- AWS Control Tower libreria di controlli
- AWS Standard di base sulle migliori pratiche di sicurezza
- Regole gestite da AWS Config
- AWS Trusted Advisor check reference
- Monitoraggio dei risultati dei AWS Trusted Advisor controlli con Amazon EventBridge
- Gestione delle AWS Config regole per tutti gli account della tua organizzazione
- AWS Config e AWS Organizations
- Rendilo disponibile AMI al pubblico per l'uso su Amazon EC2

Video correlati:

- Best Practices for securing your multi-account environment
- Approfondisci IAM Access Analyzer

SEC03-BP08 Condividi le risorse in modo sicuro all'interno della tua organizzazione

Con l'aumento del numero di carichi di lavoro, è possibile che sia necessario condividere l'accesso alle risorse in tali carichi di lavoro o eseguire il provisioning delle risorse più volte su più account. Possono esistere costrutti per segmentare il proprio l'ambiente, come ambienti di sviluppo, di test e di produzione. Tuttavia, la presenza di costrutti di separazione non limita la possibilità di condivisione sicura. La condivisione di componenti sovrapposti consente di ridurre i costi operativi e di garantire un'esperienza coerente, senza dover intuire cosa potrebbe sfuggire durante la creazione della stessa risorsa più volte.

Risultato desiderato: ridurre al minimo gli accessi involontari tramite l'uso di metodi sicuri di condivisione delle risorse all'interno dell'organizzazione e contribuire alle iniziative di prevenzione della perdita dei dati. Ridurre i costi operativi rispetto alla gestione dei singoli componenti, ridurre gli errori dovuti alla creazione manuale dello stesso componente più volte e aumentare la scalabilità dei carichi di lavoro. Si riducono i tempi di risoluzione in caso di guasti multipli e si aumenta la sicurezza nel determinare quando un componente non è più necessario. Per linee guida prescrittive sull'analisi delle risorse condivise all'esterno, consulta SEC03-BP07 Analizza l'accesso pubblico e tra account.

Anti-pattern comuni:

- Mancanza di un processo per il monitoraggio continuo e segnalazione automatica di condivisioni esterne inaspettate.
- Mancanza di una linea di base su ciò che deve e ciò che non deve essere condiviso.
- Scelta di una policy di ampia apertura piuttosto che di una condivisione esplicita quando richiesto.
- Creazione manuale di risorse fondamentali che si sovrappongono quando necessario.

Livello di rischio associato se questa best practice non fosse adottata: medio

Guida all'implementazione

Progetta controlli e modelli di accesso per gestire il consumo di risorse condivise in modo sicuro e solo con entità fidate. Monitora le risorse condivise e controllane in modo costante l'accesso, ricevendo un avviso in caso di condivisione inappropriata o inaspettata. Consulta Analisi dell'accesso

multi-account e pubblico per stabilire una governance che riduca l'accesso esterno alle sole risorse che lo richiedono e per stabilire un processo di monitoraggio continuo e avvisi automatici.

La condivisione tra account all'interno AWS Organizations è supportata da numerosi AWS servizi AWS Security Hub, come Amazon GuardDuty e AWS Backup. Questi servizi permettono di condividere i dati con un account centrale, di accedere a un account centrale o di gestire risorse e dati da un account centrale. Ad esempio, AWS Security Hub è possibile trasferire i risultati dai singoli account a un account centrale in cui è possibile visualizzare tutti i risultati. AWS Backup può eseguire un backup di una risorsa e condividerlo tra più account. Puoi usare AWS Resource Access Manager(AWS RAM) per condividere altre risorse comuni, come VPCsottoreti e allegati Transit Gateway AWS Network Firewallo Amazon pipeline. SageMaker

Per limitare il tuo account alla condivisione delle sole risorse all'interno dell'organizzazione, utilizza le policy di controllo dei servizi (SCPs) per impedire l'accesso a soggetti esterni. In caso di condivisione di risorse, combina controlli basati sull'identità e di rete per creare un perimetro di dati per l'organizzazione e proteggere la stessa da accessi involontari. Un perimetro di dati è un insieme di guardrail preventivi che aiutano a verificare che solo le identità fidate accedano a risorse fidate dalle reti previste. Questi controlli pongono limiti adeguati alle risorse condivisibili e impediscono la condivisione o l'esposizione di risorse che non sono consentite. Ad esempio, come parte del tuo perimetro di dati, puoi utilizzare le policy degli VPC endpoint e la `AWS:PrincipalOrgId` condizione per garantire che le identità che accedono ai tuoi bucket Amazon S3 appartengano alla tua organizzazione. È importante notare che SCPsnon si applicano ai ruoli o ai responsabili di servizio collegati ai servizi. AWS

Quando usi Amazon S3, disattiva il bucket Amazon S3 e IAM utilizza le policy ACLs per definire il controllo degli accessi. Per limitare l'accesso a un'origine Amazon S3 da CloudFrontAmazon, esegui la migrazione da origin access identity OAI () a origin access control OAC () che supporta funzionalità aggiuntive tra cui la crittografia lato server con. AWS Key Management Service

In alcuni casi, può essere necessario condividere le risorse al di fuori dell'organizzazione o concedere a terze parti l'accesso alle risorse stesse. Per linee guida prescrittive sulla gestione delle autorizzazioni per la condivisione esterna delle risorse, consulta Gestione delle autorizzazioni.

Passaggi dell'implementazione

1. Usa. AWS Organizations

   AWS Organizations è un servizio di gestione degli account che consente di consolidare più account Account AWS in un'organizzazione da creare e gestire centralmente. È possibile

raggruppare gli account in unità organizzative (OUs) e associare politiche diverse a ciascuna unità organizzativa per soddisfare le esigenze di budget, sicurezza e conformità. Puoi anche controllare in che modo i servizi di intelligenza AWS artificiale (AI) e machine learning (ML) possono raccogliere e archiviare dati e utilizzare la gestione multi-account dei AWS servizi integrati con Organizations.

2. Integrazione AWS Organizations con i AWS servizi.

   Quando utilizzi un AWS servizio per eseguire attività per tuo conto negli account dei membri della tua organizzazione, AWS Organizations crea un ruolo IAM collegato al servizio (SLR) per quel servizio in ogni account membro. È necessario gestire l'accesso affidabile utilizzando il AWS Management Console AWS APIs, il o il. AWS CLI Per una guida prescrittiva sull'attivazione dell'accesso affidabile, vedi Utilizzo AWS Organizations con altri AWS servizi e AWS servizi che puoi usare con Organizations.

3. Stabilisci un perimetro di dati.

   Il AWS perimetro è in genere rappresentato come un'organizzazione gestita da. AWS Organizations Oltre alle reti e ai sistemi locali, l'accesso alle AWS risorse è ciò che molti considerano il perimetro di My. AWS L'obiettivo del perimetro è verificare che l'accesso sia consentito se l'identità è attendibile, la risorsa è attendibile e la rete è conforme.

   a. Definisci e implementa i perimetri.

      Segui i passaggi descritti in Implementazione del perimetro nel white paper Building a Perimeter on per ogni condizione di autorizzazione. AWS Per linee guida prescrittive sulla protezione del livello di rete, consulta Protezione delle reti.

   b. Monitora e segnala in modo continuo.

      AWS Identity and Access Management Access Analyzer consente di identificare le risorse nell'organizzazione e negli account condivise con entità esterne. È possibile integrare IAMAccess Analyzer con AWS Security Hub per inviare e aggregare i risultati di una risorsa da IAM Access Analyzer a Security Hub per aiutare ad analizzare il livello di sicurezza del proprio ambiente. Per l'integrazione, attiva IAM Access Analyzer e Security Hub in ogni regione di ciascun account. Puoi anche utilizzarlo Regole di AWS Config per controllare la configurazione e avvisare la parte appropriata che utilizza AWS Chatbot with AWS Security Hub. A questo punto, puoi consultare i documenti di automazione AWS Systems Manager per la correzione delle risorse non conformi.

   c. Per linee guida prescrittive sul monitoraggio e l'invio di avvisi continui sulle risorse condivise a livello esterno, consulta Analisi dell'accesso multi-account e pubblico.

4. Utilizza la condivisione delle risorse nei AWS servizi e limita di conseguenza.

   Molti AWS servizi consentono di condividere risorse con un altro account o di indirizzare una risorsa in un altro account, come Amazon Machine Images (AMIs) e AWS Resource Access Manager (AWS RAM). Limita `ModifyImageAttribute` API a specificare gli account affidabili AMI con cui condividerli. Specificate la `ram:RequestedAllowsExternalPrincipals` condizione AWS RAM da utilizzare per limitare la condivisione solo alla vostra organizzazione, per impedire l'accesso da parte di identità non attendibili. Per considerazioni e linee guida prescrittive, consulta Resource sharing and external targets.

5. Utilizzalo AWS RAM per condividere in modo sicuro in un account o con altri. Account AWS

   AWS RAM ti consente di condividere in modo sicuro le risorse create con i ruoli e gli utenti del proprio account e di altri Account AWS. In un ambiente con più account, AWS RAM consente di creare una risorsa una sola volta e condividerla con altri account. Questo approccio aiuta a ridurre il sovraccarico operativo fornendo al contempo coerenza, visibilità e verificabilità attraverso le integrazioni con Amazon CloudWatch e AWS CloudTrail, che non ricevi quando utilizzi l'accesso su più account.

   Se disponi di risorse che hai condiviso in precedenza utilizzando una politica basata sulle risorse, puoi utilizzare la `PromoteResourceShareCreatedFromPolicy`API o un equivalente per promuovere la condivisione delle risorse a una condivisione di risorse completa. AWS RAM

   In alcuni casi, potrebbe essere necessario adottare ulteriori misure per condividere le risorse. Ad esempio, per condividere un'istantanea crittografata, è necessario condividere una chiave. AWS KMS

Risorse

Best practice correlate:

- SEC03-BP07 Analizza l'accesso pubblico e tra account
- SEC03-BP09 Condividi le risorse in modo sicuro con terze parti
- SEC05-BP01 Creare livelli di rete

Documenti correlati:

- Bucket owner granting cross-account permission to objects it does not own

- Come utilizzare Trust Policies con IAM
- Costruire Data Perimeter su AWS
- Come utilizzare un ID esterno per concedere a terzi l'accesso alle tue risorse AWS
- AWS servizi con cui è possibile utilizzare AWS Organizations
- Stabilire un perimetro di dati su AWS: Consenti solo alle identità affidabili di accedere ai dati aziendali

Video correlati:

- Granular Access with AWS Resource Access Manager
- Proteggi il perimetro dei dati con gli endpoint VPC
- Stabilire un perimetro di dati su AWS

Strumenti correlati:

- Esempi di policy del perimetro di dati

SEC03-BP09 Condividi le risorse in modo sicuro con terze parti

La sicurezza dell'ambiente cloud non si ferma alla tua organizzazione. L'organizzazione potrebbe affidare a terze parti la gestione di una parte dei dati. La gestione delle autorizzazioni per il sistema gestito da terze parti dovrebbe seguire la prassi di just-in-time accesso basata sul principio del privilegio minimo con credenziali temporanee. Lavorando a stretto contatto con una terza parte, puoi ridurre allo stesso momento la portata dell'impatto e il rischio di accesso non intenzionale.

Risultato desiderato: le credenziali a lungo termine AWS Identity and Access Management (IAM), le chiavi di IAM accesso e le chiavi segrete associate a un utente possono essere utilizzate da chiunque purché le credenziali siano valide e attive. L'utilizzo di un IAM ruolo e di credenziali temporanee consente di migliorare il livello di sicurezza generale riducendo lo sforzo necessario per mantenere le credenziali a lungo termine, incluso il sovraccarico di gestione e operativo di tali dati sensibili. Utilizzando un identificatore univoco universale (UUID) per l'ID esterno nella policy di IAM fiducia e mantenendo sotto controllo le IAM policy associate al IAM ruolo, è possibile controllare e verificare che l'accesso concesso alla terza parte non sia troppo permissivo. Per linee guida prescrittive sull'analisi delle risorse condivise all'esterno, consulta SEC03-BP07 Analizza l'accesso pubblico e tra account.

Anti-pattern comuni:

- Utilizzo della politica di IAM fiducia predefinita senza alcuna condizione.
- Utilizzo di IAM credenziali e chiavi di accesso a lungo termine.
- Riutilizzo esterno. IDs

Livello di rischio associato se questa best practice non fosse adottata: medio

Guida all'implementazione

Potresti voler consentire la condivisione di risorse al di fuori del tuo account AWS Organizations o concedere a terzi l'accesso al tuo account. Ad esempio, una terza parte potrebbe fornire una soluzione di monitoraggio che necessita di accedere alle risorse del tuo account. In questi casi, crea un ruolo IAM per più account con solo i privilegi necessari alla terza parte. Inoltre, definisci una policy di attendibilità utilizzando la condizione ID esterno. L'utilizzo di un ID esterno da parte tua o della terza parte può comportare la generazione di un ID univoco per ogni cliente, terza parte o tenancy. Una volta creato, l'ID univoco non deve essere controllato da nessuno, se non da te. La terza parte deve implementare un processo per collegare l'ID esterno al cliente in modo sicuro, verificabile e riproducibile.

Puoi anche utilizzare IAMRoles Anywhere per gestire i IAM ruoli per applicazioni che non rientrano in AWS tale ambito. AWS APIs

Se la terza parte non ha più bisogno di accedere al tuo ambiente, rimuovi il ruolo. Evita di fornire a terze parti credenziali a lungo termine. Mantieni la conoscenza di altri AWS servizi che supportano la condivisione. Ad esempio, AWS Well-Architected Tool consente di condividere un carico di lavoro con altri Account AWS e ti AWS Resource Access Manageraiuta a condividere in modo sicuro una AWS risorsa di tua proprietà con altri account.

Passaggi dell'implementazione

1. Utilizza ruoli multi-account per fornire l'accesso agli account esterni.

   I ruoli multi-account riducono la quantità di informazioni sensibili archiviate da account esterni e terze parti per l'assistenza ai propri clienti. I ruoli tra account diversi ti consentono di concedere l'accesso sicuro alle AWS risorse del tuo account a terze parti, ad AWS Partner esempio account o altri account dell'organizzazione, pur mantenendo la capacità di gestire e controllare tale accesso.

La terza parte può fornire il servizio da un'infrastruttura ibrida o, in alternativa, estrarre i dati in una sede esterna. IAMRoles Anywhere ti aiuta a consentire ai carichi di lavoro di terze parti di interagire in modo sicuro con i tuoi carichi di AWS lavoro e a ridurre ulteriormente la necessità di credenziali a lungo termine.

Non devi utilizzare credenziali a lungo termine o chiavi di accesso associate agli utenti per fornire accesso ad account esterni. Per fornire l'accesso multi-account invece, occorre utilizzare i ruoli multi-account.

2. Utilizza un ID esterno con le terze parti.

   L'utilizzo di un ID esterno consente di designare chi può assumere un ruolo in una politica di fiducia. IAM La policy di attendibilità può richiedere che l'utente che assume il ruolo dichiari la condizione e l'obiettivo in cui sta operando. Fornisce inoltre un modo per il proprietario dell'account di consentire che il ruolo venga assunto solo in circostanze specifiche. La funzione principale dell'ID esterno è quella di risolvere e prevenire il problema del "confused deputy" (delegato confuso).

   Utilizza un ID esterno se sei un Account AWS proprietario e hai configurato un ruolo per una terza parte che accede ad altri Account AWS oltre al tuo, o quando sei nella posizione di assumere ruoli per conto di diversi clienti. Collabora con la tua terza parte o AWS Partner stabilisci una condizione di ID esterna da includere nella politica di IAM fiducia.

3. Usa un dispositivo esterno IDs universalmente unico.

   Implementa un processo che generi un valore univoco casuale per un ID esterno, ad esempio un identificatore univoco universale (). UUID Il riutilizzo di dati esterni da parte di terzi IDs tra diversi clienti non risolve il problema della confusione secondaria, in quanto il cliente A potrebbe essere in grado di visualizzare i dati del cliente B utilizzando il ruolo ARN del cliente B insieme all'ID esterno duplicato. In un ambiente multi-tenant, in cui una terza parte supporta più clienti con clienti diversi Account AWS, la terza parte deve utilizzare un ID univoco diverso come ID esterno per ciascuno. Account AWS La terza parte è responsabile del rilevamento dei duplicati esterni IDs e della mappatura sicura di ciascun cliente al rispettivo ID esterno. La terza parte deve verificare di poter assumere il ruolo solo quando specifica l'ID esterno. La terza parte deve astenersi dall'archiviare il ruolo del cliente ARN e l'ID esterno fino a quando non sarà richiesto l'ID esterno.

   L'ID esterno non viene trattato come un segreto, ma non deve essere un valore facilmente individuabile, come un numero di telefono, un nome o un ID account. Rendi l'ID esterno un campo di sola lettura, in modo che non possa essere modificato per rappresentare la configurazione.

L'ID esterno può essere generato da te o dalla terza parte. Definisci un processo per stabilire chi è responsabile della generazione dell'ID. Indipendentemente dall'entità che crea l'ID esterno, la terza parte fa rispettare l'univocità e i formati in modo coerente tra i clienti.

4. Rendi obsolete le credenziali a lungo termine fornite dal cliente.

   Deprecate l'uso di credenziali a lungo termine e utilizzate ruoli tra account o Roles Anywhere. IAM Se devi utilizzare credenziali a lungo termine, stabilisci un piano per migrare verso l'accesso basato sui ruoli. Per i dettagli sulla gestione delle chiavi, consulta Gestione delle identità. Collabora inoltre con il tuo Account AWS team e le terze parti per stabilire un manuale di mitigazione del rischio. Per un prontuario su come rispondere e mitigare il potenziale impatto di un incidente di sicurezza, consulta Risposta agli imprevisti.

5. Verifica che la configurazione presenti indicazioni prescrittive o sia automatizzata.

   La policy creata per l'accesso multi-account ai tuoi account deve attenersi al principio del privilegio minimo. La terza parte deve fornire un documento relativo alla politica relativa al ruolo o un meccanismo di configurazione automatizzato che utilizzi un AWS CloudFormation modello o un modello equivalente. In questo modo si riduce la possibilità di errori associati alla creazione manuale della policy e si offre un audit trail. Per ulteriori informazioni sull'utilizzo di un AWS CloudFormation modello per creare ruoli tra account, consulta Ruoli tra account.

   La terza parte deve fornire un meccanismo di configurazione automatizzato e verificabile. Tuttavia, utilizzando il documento della policy sui ruoli che delinea gli accessi necessari, è possibile automatizzare l'impostazione del ruolo. Utilizzando un AWS CloudFormation modello o un modello equivalente, è necessario monitorare le modifiche con il rilevamento delle deviazioni come parte della pratica di audit.

6. Tieni conto delle modifiche.

   La struttura del tuo account, la tua necessità di una terza parte o l'offerta di servizi che ti viene fornita possono cambiare. Occorre anticipare cambiamenti e guasti, quindi pianificare di conseguenza con le persone, i processi e le tecnologie adeguati. Sottoponi periodicamente a audit il livello di accesso fornito e implementa metodi di rilevamento per avvisare l'utente di cambiamenti inattesi. Monitora e verifica l'uso del ruolo e del datastore esterno. IDs Occorre essere pronti a revocare l'accesso a terze parti, in modo temporaneo o permanente, in seguito a modifiche o modelli di accesso imprevisti. Inoltre, valuta l'impatto dell'operazione di revoca, compreso il tempo necessario per eseguirla, le persone coinvolte, il costo e l'impatto su altre risorse.

   Per linee guida prescrittive sui metodi di rilevamento, consulta le best practice di rilevamento.

Risorse

Best practice correlate:

- SEC02-BP02 Usa credenziali temporanee
- SEC03-BP05 Definisci barriere di autorizzazione per la tua organizzazione
- SEC03-BP06 Gestisci l'accesso in base al ciclo di vita
- SEC03-BP07 Analizza l'accesso pubblico e tra account
- SEC04 Rilevamento

Documenti correlati:

- Bucket owner granting cross-account permission to objects it does not own
- Come utilizzare le politiche di fiducia con IAM i ruoli
- Delega l'accesso tra diversi Account AWS ruoli IAM
- Come posso accedere alle risorse in un altro modo Account AWS ? IAM
- Le migliori pratiche di sicurezza in IAM
- Cross-account policy evaluation logic
- Come utilizzare un ID esterno per concedere l'accesso alle proprie AWS risorse a terzi
- Raccolta di informazioni dalle AWS CloudFormation risorse create in account esterni con risorse personalizzate
- Utilizzo sicuro di un ID esterno per accedere agli AWS account di proprietà di altri
- Estendi IAM i ruoli ai carichi di lavoro esterni IAM con IAM Roles Anywhere

Video correlati:

- Come posso consentire agli utenti o ai ruoli un Account AWS accesso separato al mio? Account AWS
- AWS re:Invent 2018: diventa un IAM policy master in 60 minuti o meno
- AWS Knowledge Center Live: IAM migliori pratiche e decisioni di progettazione

Esempi correlati:

- Well-Architected Lab - Assunzione di ruoli trasversali in Lambda (Level IAM 300)

- Configure cross-account access to Amazon DynamoDB
- AWS STS Strumento di interrogazione di rete

Rilevamento

Domanda

- SEC4. In che modo individui ed esamini gli eventi di sicurezza?

SEC4. In che modo individui ed esamini gli eventi di sicurezza?

Acquisisci e analizza gli eventi a partire da log e metriche per acquistare visibilità. Agisci su eventi di sicurezza e potenziali minacce per contribuire a rendere sicuro il carico di lavoro.

Best practice

- SEC04-BP01 Configurare la registrazione dei servizi e delle applicazioni
- SEC04-BP02 Acquisizione di log, risultati e metriche in posizioni standardizzate
- SEC04-BP03 Correlare e arricchire gli avvisi di sicurezza
- SEC04-BP04 Avviare la riparazione per le risorse non conformi

SEC04-BP01 Configurare la registrazione dei servizi e delle applicazioni

Mantieni i log degli eventi di sicurezza dei servizi e delle applicazioni. Si tratta di un principio fondamentale di sicurezza per audit, indagini e casi d'uso operativi e di un requisito di sicurezza comune basato su standard, politiche e procedure di governance, rischio e conformità (GRC).

Risultato desiderato: un'organizzazione dovrebbe essere in grado di recuperare in modo affidabile e coerente i registri degli eventi di sicurezza da AWS servizi e applicazioni in modo tempestivo quando necessario per adempiere a un processo o a un obbligo interno, come la risposta a un incidente di sicurezza. Considera la possibilità di centralizzare i log per migliori risultati operativi.

Anti-pattern comuni:

- Log archiviati in modo perpetuo o eliminati troppo presto.
- Tutti possono accedere ai log.
- Affidamento totale a processi manuali per la governance e l'utilizzo dei log.
- Archiviazione di ogni singolo tipo di log nel caso in cui sia necessario.

- Controllo dell'integrità del log solo quando è necessario.

Vantaggi derivanti dall'adozione di questa best practice: implementazione di un meccanismo di analisi delle cause principali (RCA) per gli incidenti di sicurezza e una fonte di prove per gli obblighi di governance, rischio e conformità.

Livello di rischio associato se questa best practice non fosse adottata: elevato

Guida all'implementazione

Durante un'indagine di sicurezza o in altri casi d'uso basati sui tuoi requisiti, devi essere in grado di esaminare i log pertinenti per registrare e comprendere l'intera portata e la tempistica dell'incidente. I log sono necessari anche per la generazione di avvisi, che indicano il verificarsi di determinate azioni di interesse. È fondamentale selezionare, attivare, memorizzare e configurare i meccanismi di query e recupero e gli avvisi.

Passaggi dell'implementazione

- Seleziona e utilizza le origini dei log. Prima di un'indagine di sicurezza, devi acquisire i log pertinenti per ricostruire in modo retroattivo l'attività in un Account AWS. Seleziona le origini dei log pertinenti per i carichi di lavoro.

  I criteri di selezione delle origini dei log devono basarsi sui casi d'uso richiesti dall'azienda. Stabilisci un percorso per ogni Account AWS utilizzo AWS CloudTrail o un AWS Organizations percorso e configura un bucket Amazon S3 per questo.

  AWS CloudTrail è un servizio di registrazione che tiene traccia delle API chiamate effettuate rispetto a un' Account AWS attività del servizio di acquisizione AWS . È attivato per impostazione predefinita con una conservazione di 90 giorni degli eventi di gestione che possono essere recuperati tramite CloudTrail la cronologia degli eventi utilizzando il, il AWS Management Console, o un. AWS CLI AWS SDK Per una conservazione e una visibilità più lunghe degli eventi relativi ai dati, crea un CloudTrail percorso e associalo a un bucket Amazon S3 e, facoltativamente, a un gruppo di log Amazon. CloudWatch In alternativa, puoi creare un CloudTrail lago che conservi i CloudTrail log per un massimo di sette anni e fornisca una funzione di interrogazione basata su una struttura di interrogazione SQL

  AWS consiglia ai clienti di VPC attivare il traffico di rete e di registrare utilizzando rispettivamente i log delle query del resolver VPCFlow DNS Logs e Amazon Route 53 e di trasmetterli in streaming su un bucket Amazon S3 o un gruppo di log. CloudWatch È possibile creare un log di flusso per

una, una VPC sottorete o un'interfaccia di rete. VPC Per VPC Flow Logs, puoi essere selettivo su come e dove utilizzare Flow Logs per ridurre i costi.

AWS CloudTrail Logs, VPC Flow Logs e i log delle query del resolver Route 53 sono le fonti di registrazione di base per supportare le indagini di sicurezza. AWS Puoi anche utilizzare Amazon Security Lake per raccogliere, normalizzare e archiviare questi dati di log in formato Apache Parquet e Open Cybersecurity Schema Framework (OCSF), pronto per l'interrogazione. Security Lake supporta anche altri AWS log e log provenienti da fonti di terze parti.

AWS i servizi possono generare log non acquisiti dalle fonti di log di base, come log di Elastic Load Balancing, log AWS WAF , registri dei AWS Config registratori, risultati di Amazon GuardDuty , registri di controllo di Amazon Elastic Kubernetes Service (Amazon) e registri delle applicazioni e del sistema operativo delle istanze EKS Amazon. EC2 Per un elenco completo delle opzioni di log e monitoraggio, consulta l'Appendice A: definizioni delle capacità del cloud, log ed eventi della AWS Security Incident Response Guide.

- Funzionalità di registrazione delle ricerche per ogni AWS servizio e applicazione: ogni servizio e applicazione offre opzioni per l'archiviazione dei log, AWS ognuna delle quali con le proprie funzionalità di conservazione e ciclo di vita. I due servizi di archiviazione dei log più comuni sono Amazon Simple Storage Service (Amazon S3) e Amazon. CloudWatch Per lunghi periodi di conservazione, è consigliabile utilizzare Amazon S3 per la sua convenienza in termini di costi e per la flessibilità del ciclo di vita. Se l'opzione di registrazione principale è Amazon CloudWatch Logs, come opzione, dovresti prendere in considerazione l'archiviazione dei log a cui accedi meno frequentemente su Amazon S3.
- Seleziona l'archiviazione dei log: la scelta dell'archiviazione dei log è in genere correlata allo strumento di query utilizzato, alle funzionalità di conservazione, alla conoscenza e ai costi. Le opzioni principali per l'archiviazione dei log sono un bucket Amazon S3 o un CloudWatch gruppo di log.

  Un bucket Amazon S3 offre a possibilità di un'archiviazione economica e duratura, con una policy opzionale per il ciclo di vita. È possibile eseguire query sui log archiviati nei bucket Amazon S3 mediante servizi come Amazon Athena.

  Un gruppo di CloudWatch log fornisce uno storage durevole e una funzione di interrogazione integrata tramite CloudWatch Logs Insights.
- Identifica la conservazione dei log appropriata: quando utilizzi un bucket o un gruppo di log Amazon S3 per archiviare i CloudWatch log, devi stabilire cicli di vita adeguati per ogni fonte di log per ottimizzare i costi di storage e recupero. In genere i clienti hanno a disposizione da tre mesi

a un anno di log per le query, con una conservazione fino a sette anni. La scelta di disponibilità e conservazione deve essere in linea con i requisiti di sicurezza e con un insieme di mandati statutari, normativi e aziendali.

- Utilizza la registrazione per ogni AWS servizio e applicazione con politiche di conservazione e ciclo di vita adeguate: per ogni AWS servizio o applicazione della tua organizzazione, consulta le linee guida specifiche sulla configurazione dei log:
  - Configura Trail AWS CloudTrail
  - Configura i log VPC di flusso
  - Configurazione di Amazon GuardDuty Finding Export
  - Configura AWS Config la registrazione
  - Configura il ACL traffico AWS WAF web
  - Configurare i registri del traffico di AWS Network Firewall rete
  - Configurazione dei log di accesso per Elastic Load Balancing
  - Configurazione del log delle query del risolutore Amazon Route 53
  - Configurare i RDS log di Amazon
  - Configurazione dei log EKS di Amazon Control Plane
  - Configura CloudWatch l'agente Amazon per EC2 istanze Amazon e server locali
- Seleziona e implementa meccanismi di interrogazione per i log: per le query di log, puoi utilizzare CloudWatch Logs Insights per i dati archiviati in gruppi di CloudWatch log e Amazon Athena e Amazon OpenSearch Service per i dati archiviati in Amazon S3. Puoi anche utilizzare strumenti di interrogazione di terze parti come un servizio di gestione delle informazioni di sicurezza e degli eventi (). SIEM

  Il processo di selezione di uno strumento di query dei log deve considerare gli aspetti relativi a persone, processi e tecnologia delle operazioni di sicurezza. Occorre scegliere uno strumento che soddisfi i requisiti operativi, aziendali e di sicurezza, accessibile e di cui sia possibile effettuare la manutenzione a lungo termine. Tieni presente che gli strumenti di query dei log funzionano in modo ottimale quando il numero di log da analizzare è mantenuto entro i limiti dello strumento. Non è raro avere più strumenti di query a causa di vincoli tecnici o di costo.

  Ad esempio, è possibile utilizzare uno strumento di gestione delle informazioni e degli eventi di sicurezza (SIEM) di terze parti per eseguire query sugli ultimi 90 giorni di dati, ma utilizzare Athena per eseguire query oltre i 90 giorni a causa del costo di inserimento dei log di un. SIEM Indipendentemente dall'implementazione, verifica che il tuo approccio riduca al minimo il numero

di strumenti necessari per ottimizzare l'efficienza operativa, soprattutto durante le indagini su un evento di sicurezza.

- Usa i log per gli avvisi: fornisce avvisi tramite diversi servizi di sicurezza: AWS
  - AWS Configmonitora e registra le configurazioni AWS delle risorse e consente di automatizzare la valutazione e la correzione rispetto alle configurazioni desiderate.
  - Amazon GuardDuty è un servizio di rilevamento delle minacce che monitora continuamente attività dannose e comportamenti non autorizzati per proteggere i tuoi Account AWS carichi di lavoro. GuardDuty acquisisce, aggrega e analizza informazioni da fonti quali eventi di AWS CloudTrail gestione e dati, DNS log, VPC Flow Logs e Amazon Audit logs. EKS GuardDuty estrae flussi di dati indipendenti direttamente da VPC Flow Logs CloudTrail, log di DNS query e Amazon. EKS Non è necessario gestire le policy del bucket Amazon S3 o modificare le modalità di raccolta e archiviazione dei log. È comunque consigliabile mantenere questi log a fini investigativi e di conformità.
  - AWS Security Hubfornisce un unico luogo in cui aggrega, organizza e dà priorità agli avvisi di sicurezza o ai risultati di più AWS servizi e prodotti opzionali di terze parti per offrirti una visione completa degli avvisi di sicurezza e dello stato di conformità.

  Esistono anche motori di generazione di avvisi personalizzati per gli avvisi di sicurezza non coperti da questi servizi o per gli avvisi specifici relativi al tuo ambiente. Per informazioni sulla creazione di questi avvisi e rilevamenti, consulta Detection in the Security Incident Response Guide. AWS

Risorse

Best practice correlate:

- SEC04-BP02 Acquisizione di log, risultati e metriche in posizioni standardizzate
- SEC07-BP04 Definire una gestione scalabile del ciclo di vita dei dati
- SEC10-BP06 Strumenti di pre-installazione

Documenti correlati:

- AWS Guida alla risposta agli incidenti di sicurezza
- Nozioni di base su Amazon Security Lake
- Guida introduttiva: Amazon CloudWatch Logs
- Soluzione dei partner per la sicurezza: registrazione e monitoraggio

Video correlati:

- AWS re:Invent 2022 - Presentazione di Amazon Security Lake

Esempi correlati:

- Assisted Log Enabler per AWS
- AWS Security Hub Esportazione storica dei risultati

Strumenti correlati:

- Snowflake for Cybersecurity

SEC04-BP02 Acquisizione di log, risultati e metriche in posizioni standardizzate

I team di sicurezza si basano su log ed esiti per analizzare gli eventi che possono indicare attività non autorizzate o modifiche non intenzionali. Per semplificare tale analisi, acquisisci i log e gli esiti di sicurezza in posizioni standardizzate. Ciò rende disponibili i punti di interesse dei dati per la correlazione e può semplificare le integrazioni degli strumenti.

Risultato desiderato: un approccio standardizzato alla raccolta, analisi e visualizzazione di dati di log, esiti e metriche. I team di sicurezza possono correlare, analizzare e visualizzare in modo efficiente i dati di sicurezza su sistemi diversi per scoprire potenziali eventi di sicurezza e identificare le anomalie. I sistemi di gestione delle informazioni e degli eventi di sicurezza (SIEM) o altri meccanismi sono integrati per interrogare e analizzare i dati di registro per risposte tempestive, tracciare e intensificare gli eventi di sicurezza.

Anti-pattern comuni:

- I team hanno e gestiscono in modo indipendente la raccolta di log e metriche che non è coerente con la strategia di registrazione dell'organizzazione.
- I team non dispongono di controlli di accesso adeguati per limitare visibilità e alterazione dei dati raccolti.
- I team non gestiscono log, esiti e metriche di sicurezza nell'ambito della loro policy di classificazione dei dati.
- I team trascurano i requisiti di sovranità e localizzazione dei dati durante la configurazione delle raccolte di dati.

Vantaggi dell'adozione di questa best practice: una soluzione di log standardizzata per raccogliere ed effettuare query su dati ed eventi dei log garantisce approfondimenti migliori ricavati dalle informazioni in essi contenute La configurazione di un ciclo di vita automatizzato per i dati di log raccolti può ridurre i costi sostenuti per l'archiviazione dei log. È possibile creare un controllo degli accessi granulare per le informazioni di log raccolte, in base a sensibilità dei dati e modelli di accesso richiesti dai team. Puoi integrare strumenti per correlare, visualizzare e ricavare informazioni dai dati.

Livello di rischio associato se questa best practice non fosse adottata: medio

Guida all'implementazione

La crescita dell' AWS utilizzo all'interno di un'organizzazione si traduce in un numero crescente di carichi di lavoro e ambienti distribuiti. Dato che ciascuno di questi carichi di lavoro e ambienti genera dati sull'attività al suo interno, l'acquisizione e l'archiviazione di questi dati a livello locale rappresenta una sfida per le operazioni di sicurezza. I team addetti alla sicurezza utilizzano strumenti come i sistemi di sicurezza delle informazioni e di gestione degli eventi (SIEM) per raccogliere dati da fonti distribuite e sottoporsi a flussi di lavoro di correlazione, analisi e risposta. Ciò richiede la gestione di una serie complessa di autorizzazioni per l'accesso alle varie fonti di dati e un sovraccarico aggiuntivo nell'esecuzione dei processi di estrazione, trasformazione e caricamento (). ETL

Per superare queste sfide, prendete in considerazione l'aggregazione di tutte le fonti pertinenti dei dati dei log di sicurezza in un account di archiviazione dei log, come descritto in Organizzazione AWS dell'ambiente utilizzando più account. Ciò comprende tutti i dati relativi alla sicurezza provenienti da carichi di lavoro e log generati dai servizi AWS , come AWS CloudTrail, AWS WAF, Elastic Load Balancing e Amazon Route 53. L'acquisizione di questi dati in posizioni standardizzate e separate Account AWS con autorizzazioni appropriate per più account offre diversi vantaggi. Questa pratica aiuta a prevenire la manomissione dei log all'interno di ambienti e carichi di lavoro compromessi, fornisce un unico punto di integrazione per strumenti aggiuntivi, oltre a offrire un modello più semplificato per la configurazione della conservazione dei dati e del ciclo di vita. Valuta gli impatti della sovranità dei dati, degli ambiti di conformità e di altre normative per determinare se sono necessarie più sedi di archiviazione di dati di sicurezza e relativi periodi di conservazione.

Per semplificare acquisizione e standardizzazione di log ed esiti, prendi in considerazione Amazon Security Lake nel tuo account Log Archive. Puoi configurare Security Lake per importare automaticamente dati da fonti comuni come Route 53 CloudTrailEKS, Amazon e VPCFlow Logs. Puoi anche configurarlo AWS Security Hub come origine dati in Security Lake, consentendoti di correlare i risultati di altri AWS servizi, come Amazon GuardDuty e Amazon Inspector, con i tuoi dati di log. Puoi anche utilizzare integrazioni di origini dati di terze parti o configurare origini dati personalizzate. Tutte le integrazioni standardizzano i dati nel formato Open Cybersecurity Schema Framework (OCSF) e

vengono archiviate in bucket Amazon S3 come file Parquet, eliminando la necessità di elaborazione. ETL

L'archiviazione dei dati di sicurezza in posizioni standardizzate offre funzionalità di analisi avanzate.AWS consiglia di distribuire strumenti per l'analisi della sicurezza che operano in un AWS ambiente in un account Security Tooling separato dal proprio account Log Archive. Questo approccio consente di implementare controlli approfonditi per proteggere l'integrità e la disponibilità dei log e del processo di gestione dei log, distinti dagli strumenti che vi accedono. Prendi in considerazione l'utilizzo di servizi, come Amazon Athena, per l'esecuzione di query su richiesta che mettono in correlazione più origini dati. Puoi anche integrare strumenti di visualizzazione, come Amazon QuickSight. Le soluzioni basate sull'intelligenza artificiale sono sempre più disponibili e possono svolgere funzioni quali la traduzione degli esiti in sintesi leggibili dall'uomo e l'interazione in linguaggio naturale. Queste soluzioni sono spesso più facilmente integrate grazie a una posizione di archiviazione di dati standardizzata per le interrogazioni.

Passaggi dell'implementazione

1. Crea gli account di archiviazione di log e Security Tooling
   a. Utilizzando AWS Organizations, crea gli account Log Archive e Security Tooling in un'unità organizzativa di sicurezza. Se utilizzi AWS Control Tower per gestire la tua organizzazione, gli account Log Archive e Security Tooling vengono creati automaticamente. Configura ruoli e autorizzazioni per l'accesso a questi account e la loro amministrazione, come richiesto.
2. Configurazione delle posizioni standardizzate dei dati di sicurezza
   a. Determina la tua strategia per la creazione di posizioni di dati di sicurezza standardizzate. Puoi raggiungere questo obiettivo attraverso opzioni come approcci comuni all'architettura dei data lake, prodotti dati di terze parti o Amazon Security Lake.AWS ti consiglia di acquisire i dati di sicurezza da quelli Regioni AWS che hai attivato per i tuoi account, anche quando non vengono utilizzati attivamente.
3. Configura la pubblicazione delle origini dati nelle tue posizioni standardizzate
   a. Identificate le fonti per i vostri dati di sicurezza e configuratele per la pubblicazione nelle vostre sedi standardizzate. Valuta le opzioni per esportare automaticamente i dati nel formato desiderato rispetto a quelle in cui è necessario sviluppare ETL i processi. Con Amazon Security Lake, puoi raccogliere dati da AWS fonti supportate e sistemi integrati di terze parti.
4. Configura gli strumenti per l'accesso alle tue posizioni standardizzate
   a. Configura strumenti come Amazon Athena QuickSight, Amazon o soluzioni di terze parti per avere l'accesso richiesto alle tue sedi standardizzate. Configura questi strumenti in modo che

operino dall'account Security Tooling con accesso in lettura trasversale all'account Log Archive, se applicabile. Crea abbonati in Amazon Security Lake così da fornire a questi strumenti l'accesso ai dati.

Risorse

Best practice correlate:

- SEC01-BP01 Separa i carichi di lavoro utilizzando gli account
- SEC07-BP04 Definire la gestione del ciclo di vita dei dati
- SEC08-BP04 Applica il controllo degli accessi
- OPS08-BP02 Analizza i registri dei carichi di lavoro

Documenti correlati:

- AWS Whitepaper: Organizzazione dell'ambiente utilizzando più account AWS
- AWS Guida prescrittiva: AWS Security Reference Architecture ()AWS SRA
- Guida prescrittiva AWS : Logging and monitoring guide for application owners

Esempi correlati:

- Aggregazione, ricerca e visualizzazione di dati di log da fonti distribuite con Amazon Athena e Amazon QuickSight
- Come visualizzare i risultati di Amazon Security Lake con Amazon QuickSight
- Genera informazioni basate sull'intelligenza artificiale per Amazon Security Lake utilizzando Amazon SageMaker Studio e Amazon Bedrock
- Identifica le anomalie di sicurezza informatica nei dati di Amazon Security Lake utilizzando Amazon SageMaker
- Inserisci, trasforma e distribuisci eventi pubblicati da Amazon Security Lake ad Amazon Service OpenSearch
- Come usare AWS Security Hub e Amazon OpenSearch Service per SIEM

Strumenti correlati:

- Amazon Security Lake

- Integrazioni con i partner di Amazon Security Lake
- Open Cybersecurity Schema Framework () OCSF
- Amazon Athena
- Amazon QuickSight
- Amazon Bedrock

SEC04-BP03 Correlare e arricchire gli avvisi di sicurezza

Un'attività imprevista può generare diversi avvisi di sicurezza da origini diverse, richiedendo un'ulteriore correlazione e arricchimento per la comprensione del contesto completo. Implementa correlazione e arricchimento automatizzati degli avvisi di sicurezza per un'identificazione e una risposta agli incidenti più accurate.

Risultato desiderato: mentre l'attività generano avvisi diversi all'interno di carichi di lavoro e ambienti, i meccanismi automatizzati correlano i dati e li arricchiscono con informazioni aggiuntive. Questa pre-elaborazione presenta un quadro più dettagliato dell'evento, che aiuta gli investigatori a determinare la criticità dell'evento e a stabilire se si tratta di un incidente che richiede una risposta formale. Questo processo riduce il carico sui team di monitoraggio e investigazione.

Anti-pattern comuni:

- Gruppi diversi di persone esaminano esiti e avvisi generati da sistemi differenti, a meno che i requisiti di separazione degli incarichi non impongano altrimenti.
- L'organizzazione convoglia tutti i dati di esiti e avvisi di sicurezza in posizioni standard, ma richiede agli investigatori di eseguire correlazioni e arricchimenti manuali.
- Ti affidi esclusivamente all'intelligence dei sistemi di rilevamento delle minacce per riferire sugli esiti e stabilire la criticità.

Vantaggi dell'adozione di questa best practice: riduzione del carico cognitivo complessivo e della preparazione manuale dei dati richiesta agli investigatori grazie a correlazione e arricchimento automatizzati degli avvisi. Questa pratica può ridurre il tempo necessario per determinare se l'evento rappresenta un incidente e avviare una risposta formale. Un contesto aggiuntivo consente inoltre di valutare con precisione la reale gravità di un evento, in quanto può essere superiore o inferiore a quanto suggerito da un avviso.

Livello di rischio associato se questa best practice non fosse adottata: basso

Guida all'implementazione

Gli avvisi di sicurezza possono provenire da diverse fonti interne, tra cui: AWS

- Servizi come Amazon GuardDuty, AmazonMacie AWS Security Hub, AmazonInspector e Network Access AWS ConfigAWS Identity and Access Management Access AnalyzerAnalyzer
- Avvisi provenienti dall'analisi automatizzata dei log di AWS servizio, infrastruttura e applicazione, ad esempio da Security Analytics for Amazon OpenSearch Service.
- Allarmi in risposta a cambiamenti nella tua attività di fatturazione provenienti da fonti come Amazon EventBridge, CloudWatch Amazon o. Budget AWS
- Fonti di terze parti come feed di intelligence sulle minacce e Security Partner Solutions di AWS Partner Network
- Contatta AWS Trust & Safety o altre fonti, come clienti o dipendenti interni.

Nella loro forma più elementare, gli avvisi contengono informazioni su chi (il principale o l'identità) sta facendo cosa (l'azione intrapresa) e cosa (le risorse interessate). Per ognuna di queste origini, individua le modalità con cui puoi creare mappature tra gli identificatori per queste identità, azioni e risorse come base per eseguire la correlazione. Ciò può avvenire sotto forma di integrazione delle fonti di allarme con uno strumento di gestione delle informazioni e degli eventi di sicurezza (SIEM) per eseguire correlazioni automatizzate, creare pipeline ed elaborazione dei dati personalizzate o una combinazione di entrambe.

Un esempio di servizio in grado di eseguire la correlazione è Amazon Detective. Detective acquisisce continuamente avvisi provenienti da varie fonti AWS e da terze parti e utilizza diverse forme di intelligence per creare un grafico visivo delle loro relazioni per facilitare le indagini.

Sebbene la criticità iniziale di un avviso sia un aiuto per la definizione delle priorità, il relativo contesto di generazione ne determina la vera criticità. Ad esempio, Amazon GuardDuty può avvisare che un'EC2istanza Amazon all'interno del tuo carico di lavoro sta eseguendo una query su un nome di dominio inaspettato. GuardDuty potrebbe assegnare una bassa criticità a questo avviso da solo. Tuttavia, la correlazione automatica con altre attività relative al momento dell'avviso potrebbe rivelare che diverse centinaia di EC2 istanze sono state distribuite dalla stessa identità, con un conseguente aumento dei costi operativi complessivi. In tal caso, GuardDuty potrebbe pubblicare il contesto dell'evento correlato come nuovo avviso di sicurezza e regolare la criticità su un livello elevato, in modo da accelerare ulteriori azioni.

Passaggi dell'implementazione

1. Identifica le origini delle informazioni sugli avvisi di sicurezza. Scopri come gli avvisi provenienti da questi sistemi rappresentano identità, azioni e risorse per determinare dove è possibile una correlazione.
2. Stabilisci un meccanismo per acquisire avvisi da diverse origini. Prendi in considerazione servizi come Security Hub e a questo CloudWatch scopo. EventBridge
3. Identifica le origini per correlazione e arricchimento dei dati. Le fonti di esempio includono CloudTrail VPC Flow Logs, Amazon Security Lake e log di infrastruttura e applicazioni.
4. Integra i tuoi avvisi con le tue origini di correlazione e arricchimento dei dati per creare contesti degli eventi di sicurezza più dettagliati e stabilire le criticità.
   a. Amazon Detective, SIEM Tooling o altre soluzioni di terze parti possono eseguire automaticamente un determinato livello di acquisizione, correlazione e arricchimento.
   b. Puoi anche utilizzare i AWS servizi per crearne di tuoi. Ad esempio, puoi richiamare una AWS Lambda funzione per eseguire una query su Amazon Athena AWS CloudTrail o su Amazon Security Lake e pubblicare i risultati su. EventBridge

Risorse

Best practice correlate:

- SEC10-BP03 Prepara le funzionalità forensi
- OPS08-BP04 Crea avvisi utilizzabili
- REL06-BP03 Invia notifiche (elaborazione e allarme in tempo reale)

Documenti correlati:

- AWS Security Incident Response Guide

Esempi correlati:

- Come arricchire i risultati con i metadati degli account AWS Security Hub
- Come usare AWS Security Hub e Amazon OpenSearch Service per SIEM

Strumenti correlati:

- Amazon Detective
- Amazon EventBridge
- AWS Lambda
- Amazon Athena

SEC04-BP04 Avviare la riparazione per le risorse non conformi

I controlli investigativi possono segnalare la presenza di risorse non conformi ai requisiti di configurazione. È possibile avviare interventi correttivi definiti in modo programmatico, sia manualmente sia automaticamente, per riparare queste risorse e ridurre al minimo gli impatti potenziali. Quando definisci le correzioni in modo programmatico, puoi intraprendere azioni rapide e coerenti.

Sebbene l'automazione possa migliorare le operazioni di sicurezza, occorre implementarla e gestirla con attenzione. Implementa meccanismi di supervisione e controllo opportuni per verificare che le risposte automatizzate siano efficaci, accurate e in linea con le policy organizzative e la propensione al rischio.

Risultato desiderato: definizione di standard di configurazione delle risorse insieme a passaggi correttivi in caso di rilevamento di una mancata conformità. Dove possibile, hai definito gli interventi correttivi in modo programmatico, in modo da avviarli manualmente o attraverso l'automazione. Sono disponibili sistemi di rilevamento per identificare le risorse non conformi e pubblicare avvisi in strumenti centralizzati monitorati dal personale di sicurezza. Questi strumenti supportano l'esecuzione degli interventi correttivi programmatici, manualmente o automaticamente. Le soluzioni automatiche dispongono di meccanismi di supervisione e controllo adeguati per regolarne l'utilizzo.

Anti-pattern comuni:

- Automazione implementata, ma non si riescono a testare e convalidare a fondo le azioni correttive. Ciò può comportare conseguenze indesiderate, come l'interruzione delle operazioni aziendali legittime o l'instabilità del sistema.
- L'automazione migliora tempi e procedure di risposta, ma senza un monitoraggio adeguato e senza meccanismi che consentano l'intervento umano e la valutazione, quando necessario.
- Ci si affida esclusivamente agli interventi correttivi, senza considerarli come parte di un programma più ampio di risposta agli incidenti e di ripristino.

Vantaggi dell'adozione di questa best practice: gli interventi correttivi automatici possono rispondere alle configurazioni errate più rapidamente rispetto ai processi manuali, il che contribuisce a ridurre al minimo i potenziali impatti aziendali e a ridurre la finestra di opportunità per usi indesiderati. Nel definire gli interventi correttivi in modo programmatico, questi vengono applicate in modo coerente, il che riduce il rischio di errore umano. L'automazione è altresì in grado di gestire un volume maggiore di avvisi contemporaneamente, il che è molto importante negli ambienti che operano su larga scala.

Livello di rischio associato se questa best practice non fosse adottata: medio

Guida all'implementazione

Come illustrato in SEC01-BP03 Identificare e convalidare gli obiettivi di controllo, servizi come AWS Config aiutano a monitorare la configurazione delle risorse nei tuoi account per verificarne la conformità ai tuoi requisiti. Quando vengono rilevate risorse non conformi, ti consigliamo di configurare l'invio di avvisi a una soluzione cloud Security Posture Management (), ad esempio per facilitare la riparazione. CSPM AWS Security Hub Queste soluzioni offrono agli investigatori della sicurezza il punto centrale per il monitoraggio dei problemi e l'adozione di misure correttive.

Mentre alcune situazioni di non conformità delle risorse sono uniche e la loro risoluzione richiede il giudizio umano, altre situazioni hanno una risposta standard che si può definire in maniera programmatica. Ad esempio, una risposta standard a un gruppo di VPC sicurezza configurato in modo errato potrebbe consistere nel rimuovere le regole non consentite e avvisare il proprietario. È possibile definire le risposte nelle funzioni di AWS Lambda, nei documenti di AWS Systems Manager Automation o tramite altri ambienti di codice di propria preferenza. Assicurati che l'ambiente sia in grado di autenticarsi per l' AWS utilizzo di un IAM ruolo con il minor numero di autorizzazioni necessario per intraprendere azioni correttive.

Una volta definita la riparazione desiderata, è possibile determinare i mezzi preferiti per avviarla. AWS Config può avviare interventi correttivi per voi. Se utilizzi Security Hub, puoi farlo tramite azioni personalizzate, che pubblicano le informazioni di ricerca su Amazon EventBridge. Una EventBridge regola può quindi avviare la riparazione. È possibile configurare l'azione personalizzata in Security Hub in modo che l'esecuzione sia automatica o manuale.

Per le azioni correttive programmatiche, ti consigliamo di disporre di log e audit completi delle azioni intraprese e dei relativi risultati. Rivedi e analizza questi log per valutare l'efficacia dei processi automatizzati e identificare le aree di miglioramento. Acquisisci i log in Amazon CloudWatch Logs e i risultati delle riparazioni sotto forma di note di ricerca in Security Hub.

Come punto di partenza, prendi in considerazione Automated Security Response on AWS, che offre soluzioni predefinite per risolvere gli errori di configurazione di sicurezza più comuni.

Passaggi dell'implementazione

1. Analizza e assegna priorità agli avvisi.
    a. Consolida gli avvisi di sicurezza provenienti da vari AWS servizi in Security Hub per visibilità, prioritizzazione e correzione centralizzate.
2. Sviluppa soluzioni correttive.
    a. Utilizza servizi come Systems Manager ed AWS Lambda esegui riparazioni programmatiche.
3. Configura le modalità di avvio delle correzioni.
    a. Utilizzando Systems Manager, definisci azioni personalizzate su cui pubblicare i risultati EventBridge. Configura queste azioni in modo l'avvio avvenga manualmente o automaticamente.
    b. Puoi anche utilizzare Amazon Simple Notification Service (SNS) per inviare notifiche e avvisi alle parti interessate (come il team di sicurezza o i team di risposta agli incidenti) per un intervento manuale o un'escalation, se necessario.
4. Rivedi e analizza i log delle correzioni per verificarne efficacia e miglioramenti.
    a. Invia l'output del log a Logs. CloudWatch Acquisisci i risultati come note sull'esito in Security Hub.

Risorse

Best practice correlate:

- SEC06-BP03 Riduci la gestione manuale e l'accesso interattivo

Documenti correlati:

- AWS Security Incident Response Guide - Detection

Esempi correlati:

- Risposta di sicurezza automatizzata attiva AWS
- Monitora le coppie di chiavi delle EC2 istanze utilizzando AWS Config
- Create AWS Config custom rules by using AWS CloudFormation Guard policies
- Correggi automaticamente istanze e cluster Amazon RDS DB non crittografati

Strumenti correlati:

- AWS Systems Manager Automation
- Risposta di sicurezza automatizzata attiva AWS

Protezione dell'infrastruttura

Questions

- SEC5. In che modo proteggi le risorse di rete?
- SEC6. In che modo proteggi le risorse di calcolo?

SEC5. In che modo proteggi le risorse di rete?

Qualsiasi carico di lavoro che abbia una qualche forma di connettività di rete, che si tratti di Internet o di una rete privata, richiede più livelli di difesa per proteggere da minacce esterne e interne basate sulla rete.

Best practice

- SEC05-BP01 Creare livelli di rete
- SEC05-BP02 Controlla il flusso di traffico all'interno dei livelli di rete
- SEC05-BP03 Implementare la protezione basata sull'ispezione
- SEC05-BP04 Automatizza la protezione della rete

SEC05-BP01 Creare livelli di rete

Segmenta la topologia di rete in diversi livelli basati su raggruppamenti logici dei componenti del carico di lavoro in base alla sensibilità dei dati e ai requisiti di accesso. Distingui tra i componenti che richiedono l'accesso in entrata da Internet, come gli endpoint Web pubblici, e quelli che necessitano solo di un accesso interno, come i database.

Risultato desiderato: i livelli della rete fanno parte di un defense-in-depth approccio integrale alla sicurezza che integra la strategia di autenticazione e autorizzazione delle identità dei carichi di lavoro. I livelli sono implementati in base alla sensibilità dei dati e ai requisiti di accesso, con meccanismi adeguati in termini di flusso e controllo del traffico.

Anti-pattern comuni:

- Crei tutte le risorse in un'unica VPC o in una sottorete.
- Creazione dei livelli di rete senza considerare i requisiti di sensibilità dei dati, il comportamento dei componenti o la loro funzionalità.
- Utilizzate VPCs le sottoreti come impostazioni predefinite per tutte le considerazioni a livello di rete e non tenete conto del modo AWS in cui i servizi gestiti influenzano la topologia.

Vantaggi dell'adozione di questa best practice: la definizione di livelli di rete è il primo passo per limitare i percorsi superflui lungo la rete, in particolare quelli che conducono a sistemi e dati critici. In tal modo gli attori non autorizzati avranno più difficoltà ad accedere alla rete e a navigare verso altre risorse al suo interno. I livelli di rete discreti riducono l'ambito di analisi dei sistemi di ispezione, ad esempio per il rilevamento delle intrusioni o la prevenzione del malware. Di conseguenza, si riduce il potenziale di falsi positivi e il sovraccarico di elaborazione non necessario.

Livello di rischio associato se questa best practice non fosse adottata: elevato

Guida all'implementazione

Quando si progetta l'architettura di un carico di lavoro, è comune separare i componenti in diversi livelli in base alle rispettive responsabilità. Ad esempio, un'applicazione Web può avere un livello di presentazione, uno di applicazione e uno di dati. È possibile adottare un approccio simile quando progetti la tua topologia di rete. I controlli di rete sottostanti possono contribuire a far rispettare i requisiti di accesso ai dati del carico di lavoro. Ad esempio, in un'architettura di applicazioni Web a tre livelli, puoi archiviare i file statici del livello di presentazione su Amazon S3 e servirli da una rete di distribuzione di contenuti CDN (), come Amazon. CloudFront Il livello dell'applicazione può avere endpoint pubblici che un Application Load Balancer ALB () serve in una sottorete pubblica VPCAmazon (simile a una zona demilitarizzata DMZ o), con servizi di back-end distribuiti in sottoreti private. Il livello dati che funge da host per risorse come database e file system condivisi può risiedere in sottoreti private diverse dalle risorse del livello applicativo. In corrispondenza dei confini di ciascuno di questi livelli (sottorete pubblicaCDN, sottorete privata), puoi implementare controlli che consentono solo al traffico autorizzato di attraversare tali confini.

Analogamente alla modellazione dei livelli di rete in base allo scopo funzionale dei componenti del carico di lavoro, occorre prendere in considerazione anche la sensibilità dei dati elaborati. Utilizzando l'esempio dell'applicazione Web, mentre tutti i servizi del carico di lavoro possono risiedere all'interno del livello di applicazione, servizi diversi possono elaborare dati con livelli di sensibilità differenti. In questo caso, la suddivisione del livello dell'applicazione utilizzando più sottoreti private, diverse tra loro o addirittura diverse VPCs Account AWS per ogni livello di sensibilità dei dati Account AWS, può essere appropriata VPCs in base ai requisiti di controllo.

Un'ulteriore considerazione per i livelli di rete consiste nella coerenza del comportamento dei componenti del carico di lavoro. Continuando con l'esempio, nel livello di applicazione possono essere presenti servizi che accettano input dagli utenti finali o integrazioni di sistemi esterni intrinsecamente più rischiosi rispetto agli input di altri servizi. A titolo esemplificativo, si possono citare il caricamento di file, l'esecuzione di script di codice, la scansione di e-mail e così via. La collocazione di questi servizi nel proprio livello di rete contribuisce a creare un limite di isolamento più forte attorno a essi e può evitare che il loro comportamento unico crei falsi positivi in termini di allarmi nei sistemi di ispezione.

Come parte della progettazione, considerate come l'uso dei servizi AWS gestiti influenzi la topologia della rete. Scopri come servizi come Amazon VPC Lattice possono aiutarti a semplificare l'interoperabilità dei componenti del carico di lavoro tra i livelli di rete. Durante l'utilizzo AWS Lambda, esegui la distribuzione nelle VPC sottoreti, a meno che non vi siano motivi specifici per non farlo. Determina dove sono VPC gli endpoint e AWS PrivateLinkpuoi semplificare l'adesione alle politiche di sicurezza che limitano l'accesso ai gateway Internet.

Passaggi dell'implementazione

1. Rivedi l'architettura del carico di lavoro. Raggruppa in modo logico componenti e servizi in base alle funzioni che svolgono, alla sensibilità dei dati elaborati e al loro comportamento.
2. Per i componenti che rispondono alle richieste provenienti da Internet, prendi in considerazione l'utilizzo di bilanciatori del carico o altri proxy per fornire endpoint pubblici. Esplora l'evoluzione dei controlli di sicurezza utilizzando servizi gestiti, come Amazon API Gateway CloudFront, Elastic Load Balancing, AWS Amplifye per ospitare endpoint pubblici.
3. Per i componenti in esecuzione in ambienti di elaborazione, come EC2 istanze Amazon, AWS Fargatecontenitori o funzioni Lambda, distribuiscili in sottoreti private basate sui tuoi gruppi sin dal primo passaggio.
4. Per AWS servizi completamente gestiti, come Amazon DynamoDB, Amazon Kinesis o SQSAmazon, prendi in considerazione l'utilizzo degli endpoint come impostazione predefinita per VPC l'accesso tramite indirizzi IP privati.

Risorse

Best practice correlate:

- REL02 Pianifica la topologia della tua rete
- PERF04-BP01 Scopri come la rete influisce sulle prestazioni

Video correlati:

- AWS re:Invent 2023 - fondamenti di rete AWS

Esempi correlati:

- VPCesempi
- Accedi alle applicazioni container in modo privato su Amazon ECS utilizzando AWS FargateAWS PrivateLink, e un Network Load Balancer
- Distribuisci contenuti statici in un bucket Amazon S3 tramite un VPC CloudFront

SEC05-BP02 Controlla il flusso di traffico all'interno dei livelli di rete

All'interno dei livelli della rete, utilizza un'ulteriore segmentazione per limitare il traffico solo ai flussi necessari per ogni carico di lavoro. In primo luogo, concentrati sul controllo del traffico tra Internet o altri sistemi esterni verso un carico di lavoro e il tuo ambiente (traffico nord-sud). Quindi, esamina i flussi tra diversi componenti e sistemi (traffico est-ovest).

Risultato desiderato: solo i flussi di rete necessari ai componenti dei tuoi carichi di lavoro possono comunicare tra loro e con i rispettivi client e con qualsiasi altro servizio da cui dipendono. La tua progettazione tiene conto di considerazioni come l'ingresso e l'uscita pubblici rispetto a quelli privati, la classificazione dei dati, le normative regionali e i requisiti di protocollo. Ove possibile, si favoriscono point-to-point i flussi rispetto al peering di rete come parte del principio della progettazione con privilegi minimi.

Anti-pattern comuni:

- Adozione di un approccio alla sicurezza della rete basato sul perimetro e controllare il flusso di traffico solo al confine dei livelli di rete.
- Si presume che tutto il traffico all'interno di un livello di rete sia autenticato e autorizzato.
- Applicazione dei controlli al traffico in ingresso o a quello in uscita, ma non a entrambi.
- Affidamento esclusivo per l'autenticazione e l'autorizzazione del traffico ai componenti del carico di lavoro e ai controlli di rete.

Vantaggi dell'adozione di questa best practice: questa pratica consente di ridurre il rischio di movimenti non autorizzati all'interno della rete e aggiunge un ulteriore livello di autorizzazione ai

carichi di lavoro. Eseguendo il controllo del flusso di traffico, è possibile limitare la portata dell'impatto di un incidente di sicurezza e velocizzare il rilevamento e la risposta.

Livello di rischio associato se questa best practice non fosse adottata: elevato

Guida all'implementazione

Sebbene i livelli di rete aiutino a stabilire i confini tra i componenti del carico di lavoro che svolgono una funzione, un livello di sensibilità dei dati e un comportamento simili, è possibile creare un livello di controllo del traffico molto più preciso utilizzando tecniche per segmentare ulteriormente i componenti all'interno di questi livelli secondo il principio del privilegio minimo. All'interno AWS, i livelli di rete sono definiti principalmente utilizzando sottoreti in base agli intervalli di indirizzi IP all'interno di Amazon. VPC I livelli possono anche essere definiti utilizzando diversiVPCs, ad esempio per raggruppare gli ambienti di microservizi per dominio aziendale. Quando si utilizzano più routingVPCs, mediate utilizzando un. AWS Transit Gateway Sebbene ciò fornisca il controllo del traffico a livello 4 (indirizzi IP e intervalli di porte) utilizzando gruppi di sicurezza e tabelle di routing, puoi ottenere un ulteriore controllo utilizzando servizi aggiuntivi AWS PrivateLink, come Amazon Route 53 Resolver DNS Firewall e. AWS Network FirewallAWS WAF

Comprendi e fai un inventario del flusso di dati e dei requisiti di comunicazione dei tuoi carichi di lavoro in termini di parti che avviano la connessione, porte, protocolli e livelli di rete. Valuta i protocolli disponibili per stabilire connessioni e trasmettere dati per selezionare quelli che soddisfano i tuoi requisiti di protezione (ad esempio, anziché). HTTPS HTTP Acquisisci questi requisiti sia ai limiti delle tue reti sia all'interno di ogni livello. Una volta identificati questi requisiti, esplora le opzioni per consentire il flusso del traffico richiesto solo in ciascun punto di connessione. Un buon punto di partenza è utilizzare i gruppi di sicurezza interniVPC, in quanto possono essere collegati a risorse che utilizzano un'interfaccia di rete elastica (ENI), come EC2 istanze Amazon, Amazon ECS task, Amazon EKS pods o database AmazonRDS. A differenza di un firewall Livello 4, un gruppo di sicurezza può avere una regola che consente il traffico da un altro gruppo di sicurezza in base al suo identificatore, riducendo al minimo gli aggiornamenti quando le risorse all'interno del gruppo cambiano nel tempo. Puoi anche filtrare il traffico utilizzando le regole in entrata e in uscita utilizzando i gruppi di sicurezza.

Quando il traffico si sposta da un punto all'altroVPCs, è comune utilizzare il VPC peering per un routing semplice o per un routing complesso. AWS Transit Gateway Questi approcci agevolano i flussi di traffico tra l'intervallo di indirizzi IP delle reti di origine e di destinazione. Tuttavia, se il tuo carico di lavoro richiede solo flussi di traffico tra componenti specifici e diversiVPCs, prendi in considerazione l'utilizzo di una connessione tramite. point-to-point AWS PrivateLink A tal fine,

individua quale servizio dovrebbe agire come produttore e quale dovrebbe agire come consumatore. Implementa un sistema di bilanciamento del carico compatibile per il produttore, attivalo di PrivateLink conseguenza e quindi accetta una richiesta di connessione da parte del consumatore. Al servizio del produttore viene quindi assegnato un indirizzo IP privato del consumatore VPC che il consumatore può utilizzare per effettuare richieste successive. Questo approccio riduce la necessità di eseguire il peer-to-peer delle reti. Includi i costi per l'elaborazione dei dati e il bilanciamento del carico come parte della valutazione PrivateLink.

Sebbene i gruppi di sicurezza PrivateLink aiutino a controllare il flusso tra i componenti dei carichi di lavoro, un'altra considerazione importante è come controllare a quali DNS domini possono accedere le risorse (se presenti). A seconda della DHCP configurazioneVPCs, puoi prendere in considerazione due diversi AWS servizi per questo scopo. La maggior parte dei clienti utilizza il DNS servizio predefinito Route 53 Resolver (chiamato anche Amazon DNS server o AmazonProvidedDNS) disponibile VPCs all'indirizzo +2 del suo CIDR intervallo. Con questo approccio, puoi creare regole DNS Firewall e associarle alle tue per determinare quali azioni intraprendere per gli elenchi di domini VPC che fornisci.

Se non stai utilizzando il risolutore Route 53 o se desideri integrare il Resolver con funzionalità di ispezione e controllo del flusso più approfondite oltre al filtro di dominio, prendi in considerazione l'implementazione di un AWS Network Firewall. Questo servizio ispeziona i singoli pacchetti utilizzando regole stateless o stateful per determinare se negare o consentire il traffico. Puoi adottare un approccio simile per filtrare il traffico Web in entrata verso i tuoi endpoint pubblici utilizzando AWS WAF. Per ulteriori indicazioni su questi servizi, vedere SEC05-BP03 Implementare la protezione basata sull'ispezione.

Passaggi dell'implementazione

1. Identifica i flussi di dati necessari tra i componenti dei tuoi carichi di lavoro.
2. Applica più controlli con un defense-in-depth approccio sia per il traffico in entrata che per quello in uscita, incluso l'uso di gruppi di sicurezza e tabelle di routing.
3. Utilizza i firewall per definire un controllo granulare sul traffico di rete in entrata, in uscita e attraverso l'utenteVPCs, come il Route 53 Resolver Firewall e. DNS AWS Network Firewall AWS WAF Prendi in considerazione l'utilizzo di AWS Firewall Manager per configurare e gestire a livello centrale le regole del firewall in tutta l'organizzazione.

Risorse

Best practice correlate:

- REL03-BP01 Scegli come segmentare il tuo carico di lavoro
- SEC09-BP02 Applica la crittografia in transito

Documenti correlati:

- Le migliori pratiche di sicurezza per i tuoi VPC
- AWS Network Optimization Tips
- Linee guida per la sicurezza di rete su AWS
- Proteggi il traffico VPC di rete in uscita nel Cloud AWS

Strumenti correlati:

- AWS Firewall Manager

Video correlati:

- AWS Transit Gateway architetture di riferimento per molti VPCs
- Accelerazione e protezione delle applicazioni con Amazon CloudFront AWS WAF e AWS Shield
- AWS re:Inforce 2023: Firewalls and where to put them

Esempi correlati:

- Lab: CloudFront per applicazioni Web

SEC05-BP03 Implementare la protezione basata sull'ispezione

Imposta i punti di ispezione del traffico tra i livelli di rete per verificare che i dati in transito corrispondano a categorie e schemi previsti. Analizza i flussi di traffico, i metadati e i modelli per identificare, rilevare e rispondere agli eventi in modo più efficace.

Risultato desiderato: ispezione e autorizzazione del traffico che attraversa i livelli di rete. Le decisioni di autorizzazione e rifiuto si basano su regole esplicite, informazioni sulle minacce e deviazioni dai comportamenti di base. Le protezioni diventano più severe man mano che il traffico si avvicina ai dati sensibili.

Anti-pattern comuni:

- Affidamento esclusivo alle regole del firewall basate su porte e protocolli. Mancato sfruttamento di sistemi intelligenti.
- Creazione di regole del firewall basate su specifici modelli di minaccia attuali, soggetti a modifiche.
- Ispezione solo del traffico che transita da una sottorete privata a una pubblica o da una sottorete pubblica a Internet.
- Mancata visione di base del traffico di rete da confrontare per individuare eventuali anomalie di comportamento.

Vantaggi dell'adozione di questa best practice: i sistemi di ispezione ti consentono di creare regole intelligenti, come consentire o negare il traffico solo in presenza di determinate condizioni all'interno dei dati di traffico. Approfitta dei set di regole gestiti AWS e dei partner, basati sulle più recenti informazioni sulle minacce, man mano che il panorama delle minacce cambia nel tempo. In questo modo si riduce l'onere di mantenere le regole e di ricercare gli indicatori di compromissione, riducendo il potenziale di falsi positivi.

Livello di rischio associato se questa best practice non fosse adottata: medio

Guida all'implementazione

Ottieni un controllo preciso sul traffico di rete stateful e stateless utilizzando AWS Network Firewall o altri firewall e sistemi di prevenzione delle intrusioni () Marketplace AWS che puoi implementare dietro un Gateway Load Balancer (IPS). GWLBAWS Network Firewall supporta le specifiche open source compatibili con Suricata per proteggere il carico di lavoro. IPS

Sia le soluzioni dei AWS Network Firewall fornitori che utilizzano un GWLB supportano diversi modelli di implementazione dell'ispezione in linea. Ad esempio, è possibile eseguire l'ispezione su VPC base individuale, centralizzarla o implementarla in un modello ibrido in cui il traffico est-ovest attraversa un'ispezione VPC e l'ingresso di Internet viene ispezionato di conseguenza. VPC VPC Un'altra considerazione è se la soluzione supporti l'unwrapping Transport Layer Security (TLS), che consente un'ispezione approfondita dei pacchetti per i flussi di traffico avviati in entrambe le direzioni. Per ulteriori informazioni e dettagli approfonditi su queste configurazioni, consulta la AWS Network Firewall Best Practice guide.

Se utilizzate soluzioni che eseguono out-of-band ispezioni, come l'analisi pcap dei dati a pacchetto provenienti da interfacce di rete che funzionano in modalità promiscua, potete configurare il mirroring del traffico. VPC Il traffico in mirroring viene conteggiato ai fini della larghezza di banda disponibile delle interfacce ed è soggetto agli stessi costi di trasferimento dati del traffico non in mirroring. È

possibile verificare se le versioni virtuali di questi dispositivi sono disponibili su Marketplace AWS, che possono supportare la distribuzione in linea dietro a. GWLB

Per i componenti che effettuano transazioni tramite protocolli HTTP basati su protocolli basati, proteggi la tua applicazione dalle minacce comuni con un firewall per applicazioni Web ()WAF. AWS WAFè un firewall per applicazioni Web che ti consente di monitorare e bloccare le richieste HTTP (S) che corrispondono alle tue regole configurabili prima di inviarle ad Amazon API Gateway CloudFront, Amazon AWS AppSync o un Application Load Balancer. Prendi in considerazione l'ispezione approfondita dei pacchetti quando valuti l'implementazione del firewall delle tue applicazioni Web, poiché alcuni richiedono l'interruzione TLS prima dell'ispezione del traffico. Per iniziare AWS WAF, puoi utilizzare Regole gestite da AWSin combinazione con le tue integrazioni partner o utilizzare le integrazioni dei partner esistenti.

Puoi gestire centralmente AWS WAF AWS Shield Advanced AWS Network Firewall, e i gruppi di VPC sicurezza Amazon in tutta la tua AWS organizzazione con AWS Firewall Manager.

Passaggi dell'implementazione

1. Determina se puoi disciplinare le regole di ispezione in modo ampio, ad esempio attraverso un'ispezioneVPC, o se hai bisogno di un approccio più granulare. VPC
2. Per soluzioni di ispezione in linea:
   a. Se lo utilizzi AWS Network Firewall, crea regole, politiche firewall e il firewall stesso. Una volta configurati questi elementi, puoi indirizzare il traffico verso l'endpoint del firewall per consentire l'ispezione.
   b. Se utilizzi un'appliance di terze parti con un Gateway Load Balancer GWLB (), distribuisci e configura l'appliance in una o più zone di disponibilità. Quindi, crea il tuo servizio endpointGWLB, l'endpoint e configura il routing per il tuo traffico.
3. Per out-of-band le soluzioni di ispezione:
   1. Attiva il mirroring VPC del traffico sulle interfacce in cui è necessario rispecchiare il traffico in entrata e in uscita. Puoi utilizzare EventBridge le regole di Amazon per richiamare una AWS Lambda funzione per attivare il mirroring del traffico sulle interfacce quando vengono create nuove risorse. Indirizza le sessioni di mirroring del traffico al Network Load Balancer davanti all'appliance che elabora il traffico.
4. Per soluzioni di traffico Web in entrata:
   a. Per configurare AWS WAF, inizia configurando una lista di controllo degli accessi Web (web). ACL Il Web ACL è una raccolta di regole con un'azione predefinita (ALLOWoDENY) elaborata

in serie che definisce il modo in cui l'utente WAF gestisce il traffico. Puoi creare regole e gruppi personalizzati o utilizzare gruppi di regole AWS gestiti nel tuo WebACL.

b. Una volta configurato ACL il Web, associalo a una AWS risorsa (come un Application Load Balancer, un API Gateway REST API o una CloudFront distribuzione) per iniziare a proteggere il traffico Web. ACL

Risorse

Documenti correlati:

- What is Traffic Mirroring?
- Implementing inline traffic inspection using third-party security appliances
- AWS Network Firewall architetture di esempio con routing
- Architettura di ispezione centralizzata con AWS Gateway Load Balancer e AWS Transit Gateway

Esempi correlati:

- Best practices for deploying Gateway Load Balancer
- TLSconfigurazione di ispezione per il traffico in uscita crittografato e AWS Network Firewall

Strumenti correlati:

- Marketplace AWS IDS/IPS

SEC05-BP04 Automatizza la protezione della rete

Automatizza l'implementazione delle protezioni di rete utilizzando DevOps pratiche come infrastructure as code (IaC) e pipeline CI/CD. Queste pratiche possono aiutare a tenere traccia delle modifiche apportate alle protezioni di rete attraverso un sistema di controllo delle versioni, a ridurre i tempi di implementazione delle modifiche e a rilevare se le protezioni di rete si allontanano dalla configurazione desiderata.

Risultato desiderato: definizione delle protezioni di rete con modelli e relativo inserimento in un sistema di controllo delle versioni. In caso di nuove modifiche, vengono avviate pipeline automatiche che ne orchestrano test e implementazione. I controlli delle policy e altri test statici sono in atto per convalidare le modifiche prima dell'implementazione. L'implementazione delle modifiche

avviene in un ambiente di staging per convalidare il funzionamento previsto dei controlli. Anche l'implementazione negli ambienti di produzione avviene in automatico una volta approvati i controlli.

Anti-pattern comuni:

- Affidamento ai singoli team del carico di lavoro la definizione dell'intero stack di rete, delle protezioni e delle automazioni. Mancata pubblicazione degli aspetti standard dello stack di rete e delle protezioni in modo centralizzato per consentire ai team del carico di lavoro di utilizzarli.
- Affidamento a un team di rete centrale per definire tutti gli aspetti della rete, delle protezioni e delle automazioni. Mancata delega degli aspetti specifici del carico di lavoro dello stack di rete e delle protezioni al team di quel carico di lavoro.
- Individuazione del giusto equilibrio tra centralizzazione e delega tra un team di rete e i team del carico di lavoro, ma mancata applicazione di standard di test e implementazione coerenti nei modelli IaC e nelle pipeline CI/CD. Mancata acquisizione delle configurazioni richieste negli strumenti che controllano l'aderenza dei modelli.

Vantaggi dell'adozione di questa best practice: l'utilizzo di modelli per definire le protezioni di rete consente di tracciare le modifiche e confrontarle nel tempo con un sistema di controllo delle versioni. L'uso dell'automazione per testare e implementare le modifiche crea standardizzazione e prevedibilità, aumentando le possibilità di una corretta implementazione e riducendo le configurazioni manuali ripetitive.

Livello di rischio associato se questa best practice non fosse adottata: medio

Guida all'implementazione

Una serie di controlli di protezione della rete descritti in SEC05-BP02 Controllo dei flussi di traffico all'interno dei livelli di rete e SEC 05-BP03 Implementazione della protezione basata sull'ispezione sono inclusi sistemi di regole gestite che possono essere aggiornati automaticamente in base alle più recenti informazioni sulle minacce. Esempi di protezione degli endpoint Web includono regole gestite e mitigazione automatica a livello di applicazione.AWS WAFAWS Shield Advanced DDoS Utilizza i gruppi di regole AWS Network Firewall gestite per rimanere aggiornato sugli elenchi di domini con scarsa reputazione e sulle firme delle minacce.

Oltre alle regole gestite, ti consigliamo di utilizzare DevOps procedure per automatizzare la distribuzione delle risorse di rete, delle protezioni e delle regole specificate. Puoi acquisire queste definizioni in AWS CloudFormation o in un altro strumento Infrastructure as Code (IaC) di tua scelta, trasferirle in un sistema di controllo delle versioni e implementarle mediante pipeline CI/CD. Utilizzate

questo approccio DevOps per ottenere i vantaggi tradizionali della gestione dei controlli di rete, come rilasci più prevedibili, test automatizzati con strumenti come AWS CloudFormation Guarde rilevamento degli scostamenti tra l'ambiente distribuito e la configurazione desiderata.

In base alle decisioni prese nell'ambito di SEC05-BP01 Create network layer, potreste avere un approccio di gestione centralizzato alla creazione VPCs dedicato ai flussi di ingresso, uscita e ispezione. Come descritto nella AWS Security Reference Architecture (AWS SRA), è possibile definirli VPCs in un account dedicato all'infrastruttura di rete. È possibile utilizzare tecniche simili per definire centralmente l'VPCsutilizzo dei carichi di lavoro in altri account, i relativi gruppi di sicurezza, le AWS Network Firewall distribuzioni, le regole di Route 53 Resolver e le configurazioni del DNS firewall e altre risorse di rete. Puoi condividere queste risorse con gli altri tuoi account con AWS Resource Access Manager. Grazie a questo approccio, puoi semplificare test e implementazione automatici dei controlli di rete nell'account di rete, con una sola destinazione da gestire. Puoi farlo in un modello ibrido, in cui distribuisci e condividi determinati controlli centralmente e deleghi altri controlli ai singoli team del carico di lavoro e ai rispettivi account.

Passaggi dell'implementazione

1. Stabilisci quali aspetti della rete e delle protezioni sono definiti a livello centrale e quali possono essere gestiti dai tuoi team del carico di lavoro.
2. Crea ambienti per testare e implementare le modifiche alla tua rete e alle relative protezioni. Ad esempio, utilizza un account Network Testing e uno Network Production.
3. Determina come archivierai e manterrai i tuoi modelli in un sistema di controllo delle versioni. Archivia i modelli centrali in un repository distinto da quello dei carichi di lavoro, mentre i modelli dei carichi di lavoro possono essere archiviati in repository specifici per quel carico di lavoro.
4. Crea pipeline CI/CD per testare e implementare modelli. Definisci i test per verificare che non ci siano configurazioni errate e che i modelli siano conformi agli standard aziendali.

Risorse

Best practice correlate:

- SEC01-BP06 Automatizza l'implementazione dei controlli di sicurezza standard

Documenti correlati:

- AWS Security Reference Architecture - Network account

Esempi correlati:

- AWS Deployment Pipeline Reference Architecture
- NetDevSecOpsper modernizzare le AWS implementazioni di rete
- Integrazione di test e report AWS CloudFormation di sicurezza AWS Security HubAWS CodeBuild

Strumenti correlati:

- AWS CloudFormation
- AWS CloudFormation Guard
- cfn_nag

SEC6. In che modo proteggi le risorse di calcolo?

Le risorse di calcolo nel carico di lavoro richiedono più livelli di difesa per contribuire alla protezione da minacce esterne e interne. Le risorse di calcolo includono EC2 istanze, contenitori, AWS Lambda funzioni, servizi di database, dispositivi IoT e altro ancora.

Best practice

- SEC06-BP01 Eseguire la gestione delle vulnerabilità
- SEC06-BP02 Fornitura di dati di calcolo a partire da immagini rinforzate
- SEC06-BP03 Ridurre la gestione manuale e l'accesso interattivo
- SEC06-BP04 Convalida l'integrità del software
- SEC06-BP05 Automatizza la protezione dell'elaborazione

SEC06-BP01 Eseguire la gestione delle vulnerabilità

Scansiona e correggi di frequente le vulnerabilità del codice, delle dipendenze e dell'infrastruttura per proteggerti da nuove minacce.

Risultato desiderato: creazione e gestione di un programma di gestione delle vulnerabilità. Scansiona e applica patch regolarmente a risorse come EC2 istanze Amazon, contenitori Amazon Elastic Container Service (AmazonECS) e carichi di lavoro Amazon Elastic Kubernetes Service (Amazon). EKS Configura le finestre di manutenzione per le risorse AWS gestite, come i database Amazon Relational Database Service (RDSAmazon). Utilizza la scansione statica del codice per ispezionare il codice sorgente delle applicazioni alla ricerca di problemi comuni. Prendi in considerazione la

possibilità di effettuare test di penetrazione delle applicazioni Web se l'organizzazione dispone delle competenze necessarie o se può avvalersi di un'assistenza esterna.

Anti-pattern comuni:

- Assenza di un programma di gestione delle vulnerabilità.
- Esecuzione di patch di sistema senza considerare gravità o prevenzione del rischio.
- Utilizzo di software che ha superato la data di fine del ciclo di vita () EOL fornita dal fornitore.
- Implementazione del codice in produzione prima di aver analizzato i problemi di sicurezza.

Livello di rischio associato se questa best practice non fosse adottata: elevato

Guida all'implementazione

Un programma di gestione delle vulnerabilità comprende la valutazione della sicurezza, l'identificazione dei problemi, la definizione delle priorità e l'esecuzione di operazioni di patch per risolvere i problemi. L'automazione è la chiave per la scansione continua dei carichi di lavoro alla ricerca di problemi e di esposizioni di rete non intenzionali e per l'esecuzione di interventi correttivi. L'automazione della creazione e dell'aggiornamento delle risorse fa risparmiare tempo e riduce il rischio di ulteriori problemi causati dagli errori di configurazione. Un programma di gestione delle vulnerabilità ben progettato dovrebbe tenere anche conto della verifica delle vulnerabilità durante le fasi di sviluppo e implementazione del ciclo di vita del software. L'implementazione della gestione delle vulnerabilità durante lo sviluppo e l'implementazione aiuta a ridurre le possibilità di diffusione di una vulnerabilità nell'ambiente di produzione.

L'implementazione di un programma di gestione delle vulnerabilità richiede una buona conoscenza del modello di responsabilità condivisa AWS e della sua relazione con i carichi di lavoro specifici. In base al modello di responsabilità condivisa, AWS è responsabile della protezione dell'infrastruttura di. Cloud AWS Questa infrastruttura è composta da hardware, software, reti e strutture che eseguono Cloud AWS i servizi. Sei responsabile della sicurezza nel cloud, ad esempio dei dati effettivi, della configurazione di sicurezza e delle attività di gestione delle EC2 istanze Amazon e della verifica che i tuoi oggetti Amazon S3 siano classificati e configurati correttamente. L'approccio alla gestione delle vulnerabilità può variare anche in base ai servizi utilizzati. Ad esempio, AWS gestisce l'applicazione di patch per il nostro servizio di database relazionali gestitiRDS, Amazon, ma tu sarai responsabile dell'applicazione delle patch ai database ospitati autonomamente.

AWS dispone di una gamma di servizi per aiutarti con il tuo programma di gestione delle vulnerabilità. Amazon Inspector analizza continuamente i AWS carichi di lavoro alla ricerca di problemi software

e accessi involontari alla rete. AWS Systems Manager Patch Manager aiuta a gestire l'applicazione di patch tra le EC2 istanze Amazon. È possibile visualizzare Amazon Inspector e Systems Manager in AWS Security Hub, un servizio di gestione delle posture di sicurezza nel cloud che aiuta ad automatizzare i controlli di sicurezza e a centralizzare gli AWS avvisi di sicurezza.

Amazon CodeGuru può aiutare a identificare potenziali problemi nelle applicazioni Java e Python utilizzando l'analisi statica del codice.

Passaggi dell'implementazione

- Configura Amazon Inspector: Amazon Inspector rileva automaticamente le istanze EC2 Amazon appena lanciate, le funzioni Lambda e le immagini dei container idonee inviate ad ECR Amazon e le analizza immediatamente per individuare problemi software, potenziali difetti ed esposizione involontaria della rete.
- Esegui la scansione del codice sorgente: esegui la scansione di librerie e dipendenze per individuare problemi e difetti. Amazon CodeGuru può analizzare e fornire consigli per risolvere problemi di sicurezza comuni per le applicazioni Java e Python. La OWASP Foundation pubblica un elenco di strumenti di analisi del codice sorgente (noti anche come SAST strumenti).
- Implementa un meccanismo di scansione e applicazione delle patch all'ambiente esistente, oltre a eseguire la scansione nell'ambito di un processo di creazione di una pipeline CI/CD: implementa un meccanismo di scansione e applicazione di patch per eventuali problemi nelle dipendenze e nei sistemi operativi, così proteggerli da nuove minacce. Tale processo deve essere eseguito con regolarità. La gestione delle vulnerabilità del software è essenziale per capire dove è necessario applicare le patch o risolvere i problemi del software. Stabilisci le priorità per la correzione di potenziali problemi di sicurezza incorporando le valutazioni di vulnerabilità nelle fasi iniziali della pipeline di integrazione continua/consegna continua (CI/CD). Il tuo approccio può variare in base ai AWS servizi che utilizzi. Per verificare la presenza di potenziali problemi nel software in esecuzione su EC2 istanze Amazon, aggiungi Amazon Inspector alla tua pipeline per avvisarti e interrompere il processo di creazione se vengono rilevati problemi o potenziali difetti. Amazon Inspector monitora in modo continuo le risorse. Puoi anche utilizzare prodotti open source come OWASPDependency-Check, Snyk, Open, gestori di pacchetti e strumenti per la gestione delle VAS vulnerabilità. AWS Partner
- Utilizzo AWS Systems Manager: sei responsabile della gestione delle patch per AWS le tue risorse, tra cui istanze Amazon Elastic Compute Cloud (AmazonEC2), Amazon Machine Images (AMIs) e altre risorse di elaborazione. AWS Systems Manager Patch Manager automatizza il processo di applicazione di patch alle istanze gestite con aggiornamenti correlati alla sicurezza e di altro tipo. Patch Manager può essere utilizzato per applicare patch su EC2 istanze Amazon per sistemi

operativi e applicazioni, tra cui applicazioni Microsoft, service pack di Windows e aggiornamenti di versioni minori per istanze basate su Linux. Oltre ad AmazonEC2, Patch Manager può essere utilizzato anche per applicare patch ai server locali.

Per un elenco dei sistemi operativi supportati, consulta Sistemi operativi supportati nella Guida per l'utente di Systems Manager. Puoi analizzare le istanze per visualizzare solo un report delle patch mancanti, oppure analizzare e installare automaticamente tutte le patch mancanti.

- Utilizzo AWS Security Hub: Security Hub fornisce una visione completa dello stato di sicurezza in AWS. Raccoglie dati di sicurezza su più AWS servizi e fornisce tali risultati in un formato standardizzato, che consente di dare priorità ai risultati di sicurezza tra i servizi. AWS
- Utilizza AWS CloudFormation: AWS CloudFormation è un servizio Infrastructure as Code (IaC) utile nella gestione delle vulnerabilità che automatizza l'implementazione delle risorse e standardizza l'architettura delle risorse su più account e ambienti.

Risorse

Documenti correlati:

- AWS Systems Manager
- Panoramica sulla sicurezza di AWS Lambda
- Amazon CodeGuru
- Improved, Automated Vulnerability Management for Cloud Workloads with a New Amazon Inspector
- Automatizza la gestione e la correzione delle vulnerabilità utilizzando Amazon AWS Inspector e — Parte 1 AWS Systems Manager

Video correlati:

- Securing Serverless and Container Services
- Best practice di sicurezza per il servizio di metadati delle EC2 istanze Amazon

SEC06-BP02 Fornitura di dati di calcolo a partire da immagini rinforzate

Riduci le opportunità di accesso involontario agli ambienti di runtime implementandoli da immagini rafforzate. Acquisisci dipendenze di runtime, come immagini di container e librerie di applicazioni,

solo da registri affidabili e verifica le loro firme. Crea i tuoi registri privati per archiviare immagini e librerie attendibili da utilizzare nei tuoi processi di compilazione e implementazione.

Risultato desiderato: l'allocazione delle risorse di calcolo avviene a partire da immagini di base rinforzate. Le dipendenze esterne, ad esempio immagini dei container e librerie di applicazioni, vengono recuperate solo da registri attendibili e ne vengono verificate le firme. Queste sono archiviate in registri privati a cui i processi di compilazione e implementazione possono fare riferimento. Scansiona e aggiorna con regolarità immagini e dipendenze per proteggerti da eventuali vulnerabilità scoperte di recente.

Anti-pattern comuni:

- Acquisizione di immagini e librerie da registri attendibili, ma senza verificarne la firma o eseguire scansioni delle vulnerabilità prima di metterle in uso.
- Rafforzamento delle immagini, ma senza test regolari per individuare nuove vulnerabilità o aggiornarle alla versione più recente.
- Installazione o non rimozione di pacchetti software non necessari durante il ciclo di vita previsto dell'immagine.
- Affidamento esclusivo alle patch per mantenere aggiornate le risorse di calcolo di produzione. La sola applicazione di patch può comunque far sì che nel tempo le risorse di calcolo si allontanino dallo standard rafforzato. L'applicazione delle patch può inoltre non essere in grado di rimuovere le minacce informatiche che un attore pericoloso potrebbero aver installato durante un evento di sicurezza.

Vantaggi dell'adozione di questa best practice: il rafforzamento delle immagini favorisce la riduzione del numero di percorsi disponibili nell'ambiente di runtime, che possono consentire l'accesso non intenzionale a utenti o servizi non autorizzati. Inoltre, può ridurre l'ambito dell'impatto in caso di accesso involontario.

Livello di rischio associato se questa best practice non fosse adottata: elevato

Guida all'implementazione

Per rafforzare i tuoi sistemi, occorre partire dalle versioni più recenti dei sistemi operativi, delle immagini dei container e delle librerie delle applicazioni. Applica le patch ai problemi noti. Riduci al minimo il sistema rimuovendo applicazioni, servizi, driver dei dispositivi, utenti predefiniti e altre credenziali non necessari. Adotta qualsiasi altra azione necessaria, come la disabilitazione delle

porte, per creare un ambiente che disponga solo delle risorse e delle capacità necessarie per i carichi di lavoro. Da questa linea di base è possibile installare software, agenti o altri processi necessari per scopi quali il monitoraggio del carico di lavoro o la gestione delle vulnerabilità.

È possibile ridurre l'onere del rafforzamento dei sistemi utilizzando le linee guida fornite da fonti attendibili, come le guide tecniche per l'implementazione della sicurezza del Center for Internet Security (CIS) e della Defense Information Systems Agency (DISA). STIGs Ti consigliamo di iniziare con una Amazon Machine Image (AMI) pubblicata da AWS o da un APN partner e utilizzare AWS EC2Image Builder per automatizzare la configurazione in base a una combinazione appropriata di CIS controlli e. STIG

Sebbene siano disponibili immagini rinforzate e ricette di EC2 Image Builder che applicano CIS i consigli DISA STIG o, è possibile che la loro configurazione impedisca il corretto funzionamento del software. In questa situazione, è possibile partire da un'immagine di base non protetta, installare il software e quindi applicare i CIS controlli in modo incrementale per testarne l'impatto. Per qualsiasi CIS controllo che impedisca l'esecuzione del software, verifica se invece riesci a implementare i consigli più dettagliati sulla protezione avanzata in un. DISA Tieni traccia dei diversi CIS controlli e DISA STIG configurazioni che riesci ad applicare con successo. Utilizzateli per definire di conseguenza le vostre ricette di rafforzamento delle EC2 immagini in Image Builder.

Per i carichi di lavoro containerizzati, le immagini rinforzate di Docker sono disponibili nell'archivio pubblico Amazon Elastic Container Registry (). ECR È possibile utilizzare EC2 Image Builder per rafforzare le immagini dei contenitori. AMIs

Analogamente ai sistemi operativi e alle immagini dei contenitori, è possibile ottenere pacchetti di codice (o librerie) da archivi pubblici, tramite strumenti come pip, npm, Maven e. NuGet Ti consigliamo di gestire i pacchetti di codice integrando repository privati, ad esempio all'interno di AWS CodeArtifact, con repository pubblici affidabili. Questa integrazione può gestire il recupero, l'archiviazione e la conservazione dei pacchetti per te. up-to-date I processi di creazione delle applicazioni possono quindi ottenere e testare la versione più recente di questi pacchetti insieme all'applicazione, utilizzando tecniche come Software Composition Analysis (SCA), Static Application Security Testing (SAST) e Dynamic Application Security Testing (DAST).

Per i carichi di lavoro serverless che utilizzano AWS Lambda, semplifica la gestione delle dipendenze dei pacchetti utilizzando i livelli Lambda. Usa i livelli Lambda per configurare un set di dipendenze standard condivise tra diverse funzioni in un archivio autonomo. È possibile creare e gestire i livelli tramite il relativo processo di compilazione, in modo da garantire la permanenza delle funzioni in modo centralizzato. up-to-date

Passaggi dell'implementazione

- Rafforzamento del sistema operativo. Utilizzate immagini di base provenienti da fonti attendibili come base per costruire il vostro hardenedAMIs. Usa EC2Image Builder per personalizzare il software installato sulle tue immagini.
- Rafforzamento delle risorse containerizzate. Configura le risorse containerizzate in modo che rispettino le best practice in materia di sicurezza. Quando utilizzi i contenitori, implementa la scansione delle ECR immagini nella tua pipeline di creazione e, su base regolare, nel tuo archivio di immagini da cercare CVEs nei contenitori.
- Quando si utilizza l'implementazione serverless con AWS Lambda, utilizza i livelli Lambda per separare il codice delle funzioni dell'applicazione e le librerie dipendenti condivise. Configura la firma del codice per Lambda così da garantire l'esecuzione del solo codice attendibile nelle funzioni Lambda.

Risorse

Best practice correlate:

- OPS05-BP05 Esegui la gestione delle patch

Video correlati:

- Approfondimento sulla sicurezza AWS Lambda

Esempi correlati:

- STIGCompatibile con la compilazione rapida con Image AMI Builder EC2
- Building better container images
- Using Lambda layers to simplify your development process
- Sviluppa e distribuisci AWS Lambda livelli utilizzando Serverless Framework
- Creazione di una pipeline end-to-end AWS DevSecOps CI/CD con strumenti e software open source SCA SAST DAST

SEC06-BP03 Ridurre la gestione manuale e l'accesso interattivo

Utilizza l'automazione per eseguire attività di implementazione, configurazione, manutenzione e investigazione, laddove possibile. Quando l'automazione non è disponibile, considera l'accesso manuale alle risorse di calcolo in caso di procedure di emergenza o in ambienti sicuri (sandbox).

Risultato desiderato: acquisizione mediante script programmatici e documenti di automazione (runbook) delle azioni autorizzate sulle tue risorse di calcolo. Questi runbook vengono avviati in automatico, attraverso i sistemi di rilevamento delle modifiche, o manualmente, quando è necessario il giudizio umano. L'accesso diretto alle risorse di calcolo è disponibile solo in situazioni di emergenza, quando l'automazione non è disponibile. Tutte le attività manuali vengono inserite in un log e in un processo di revisione per migliorare in modo continuo le capacità di automazione.

Anti-pattern comuni:

- Accesso interattivo alle EC2 istanze Amazon con protocolli come SSH oRDP.
- Mantenimento degli accessi dei singoli utenti, come `/etc/passwd` o gli utenti locali di Windows.
- Condivisione di una password o chiave privata per accedere a un'istanza tra più utenti.
- Installazione del software e creazione o aggiornamento manuali dei file di configurazione.
- Aggiornamento o applicazione di patch manuale al software.
- Accesso a un'istanza per risolvere i problemi.

Vantaggi dell'adozione di questa best practice: l'esecuzione di azioni automatizzate favorisce la riduzione del rischio operativo legato a modifiche non intenzionali ed errori di configurazione. La rimozione dell'uso di Secure Shell (SSH) e Remote Desktop Protocol (RDP) per l'accesso interattivo riduce l'ambito di accesso alle risorse di elaborazione. In tal modo si elimina un percorso comune per le azioni non autorizzate. Acquisire le attività di gestione delle risorse di calcolo in documenti di automazione e script di programmazione significa definire e sottoporre ad audit l'intero ambito delle attività autorizzate a un livello di dettaglio granulare.

Livello di rischio associato se questa best practice non fosse adottata: medio

Guida all'implementazione

L'accesso a un'istanza è un approccio classico all'amministrazione del sistema. Dopo aver installato il sistema operativo del server, gli utenti in genere accedono manualmente per configurare il sistema e installare il software desiderato. Nel corso del ciclo di vita del server, gli utenti possono accedere

per eseguire aggiornamenti del software, applicare patch, modificare le configurazioni e risolvere i problemi.

L'accesso manuale comporta tuttavia una serie di rischi. Richiede un server in grado di ascoltare le richieste, ad esempio un RDP servizio SSH or, in grado di fornire un potenziale percorso di accesso non autorizzato. Inoltre, aumenta il rischio di errore umano associato all'esecuzione di operazioni manuali. Le conseguenze possono essere incidenti sul carico di lavoro, danneggiamento o distruzione dei dati o altri problemi di sicurezza. L'accesso umano richiede inoltre protezioni contro la condivisione delle credenziali, creando ulteriori costi di gestione.

Per mitigare questi rischi, è possibile implementare una soluzione di accesso remoto basata su agenti, come Systems Manager AWS . AWS Systems Manager Agent (SSMAgent) avvia un canale crittografato e quindi non si basa sull'ascolto di richieste avviate dall'esterno. Prendi in considerazione la possibilità di configurare SSM Agent per stabilire questo canale su un endpoint. VPC

Systems Manager offre un controllo granulare delle modalità di interazione con le istanze gestite. Sei tu a definire le automazioni da eseguire, chi può eseguirle e quando possono essere eseguite. Systems Manager è in grado di applicare patch, installare software e apportare modifiche alla configurazione senza accesso interattivo all'istanza. Systems Manager può anche fornire l'accesso a una shell remota e registrare ogni comando richiamato e il relativo output durante la sessione nei log e in Amazon S3. AWS CloudTrailregistra le chiamate di Systems Manager APIs per l'ispezione.

Passaggi dell'implementazione

1. Installa AWS Systems Manager Agent (SSMAgent) sulle tue EC2 istanze Amazon. Verifica se SSM Agent è incluso e avviato automaticamente come parte della AMI configurazione di base.
2. Verifica che i IAM ruoli associati ai profili delle tue EC2 istanze includano la IAMpolicy `AmazonSSMManagedInstanceCore` gestita.
3. SSHRDPDisabilita e altri servizi di accesso remoto in esecuzione sulle tue istanze. Puoi farlo eseguendo script configurati nella sezione dei dati utente dei tuoi modelli di lancio o creandone di personalizzati AMIs con strumenti come EC2 Image Builder.
4. Verificate che le regole di ingresso dei gruppi di sicurezza applicabili alle vostre EC2 istanze non consentano l'accesso alla porta 22/tcp () o alla porta 3389/tcp ()SSH. RDP Implementa il rilevamento e l'invio di avvisi su gruppi di sicurezza non configurati correttamente utilizzando servizi come AWS Config.
5. Definisci automazioni, runbook ed esegui comandi appropriati in Systems Manager. Utilizzate IAM le politiche per definire chi può eseguire queste azioni e le condizioni in base alle quali sono consentite. Testa in modo approfondito queste automazioni in un ambiente non di produzione.

Richiama queste automazioni quando necessario, invece di accedere in modo interattivo all'istanza.

6. Utilizza AWS Systems Manager Session Manager per fornire un accesso interattivo alle istanze, quando necessario. Attiva la registrazione delle attività delle sessioni per mantenere un audit trail in Amazon CloudWatch Logs o Amazon S3.

Risorse

Best practice correlate:

- REL08-BP04 Esegui la distribuzione utilizzando un'infrastruttura immutabile

Esempi correlati:

- Sostituzione SSH dell'accesso per ridurre il sovraccarico di gestione e sicurezza con AWS Systems Manager

Strumenti correlati:

- AWS Systems Manager

Video correlati:

- Controllo dell'accesso della sessione utente alle istanze in AWS Systems Manager Session Manager

SEC06-BP04 Convalida l'integrità del software

Utilizza la verifica crittografica per convalidare l'integrità degli artefatti software (comprese le immagini) utilizzati dal tuo carico di lavoro. La firma crittografica del software è una tutela contro le modifiche non autorizzate eseguite negli ambienti di calcolo.

Risultato desiderato: ottenimento di tutti gli artefatti da fonti attendibili. I certificati del sito Web del fornitore sono convalidati. Gli artefatti scaricati vengono verificati a livello crittografico tramite le relative firme. Il tuo software è firmato e verificato a livello crittografico dai tuoi ambienti di elaborazione.

Anti-pattern comuni:

- Affidarsi a siti Web di fornitori attendibili per ottenere artefatti software, ma ignorare gli avvisi di scadenza dei certificati. Download senza confermare la validità dei certificati.
- Convalida dei certificati dei siti Web dei fornitori, ma senza verificare a livello crittografico gli artefatti scaricati da questi siti Web.
- Affidarsi esclusivamente a digest o hash per convalidare l'integrità del software. Gli hash stabiliscono che gli artefatti non sono stati modificati rispetto alla versione originale, ma non ne convalidano l'origine.
- Mancata firma di software, codice o librerie di proprietà, anche se utilizzati solo per le proprie implementazioni.

Vantaggi dell'adozione di questa best practice: la convalida dell'integrità degli artefatti da cui dipende il carico di lavoro consente di prevenire l'ingresso di malware negli ambienti di calcolo. La firma del software aiuta a proteggerti dall'esecuzione non autorizzata nei tuoi ambienti di calcolo. Proteggi la catena di approvvigionamento del software firmando e verificando il codice.

Livello di rischio associato se questa best practice non fosse adottata: medio

Guida all'implementazione

Immagini del sistema operativo, immagini dei container e artefatti del codice sono spesso distribuiti con controlli di integrità disponibili, ad esempio attraverso un digest o un hash. Questi permettono ai client di verificare l'integrità elaborando il proprio hash del payload e verificando che sia uguale a quello pubblicato. Sebbene questi controlli aiutino a verificare l'assenza di manomissioni del payload, non ne convalidano la provenienza dalla fonte originale (la sua provenienza). La verifica della provenienza richiede un certificato rilasciato da un'autorità attendibile per firmare digitalmente l'artefatto.

Se utilizzi un software o artefatti scaricati nel tuo carico di lavoro, controlla se il fornitore offre una chiave pubblica per la verifica della firma digitale. Ecco alcuni esempi di come AWS fornisce una chiave pubblica e le istruzioni di verifica per il software che pubblichiamo:

- EC2Image Builder: verifica la firma del download di installazione AWS TOE
- AWS Systems Manager: verifica della firma dell'agente SSM
- Amazon CloudWatch: verifica della firma del pacco dell' CloudWatch agente

Incorpora la verifica della firma digitale nei processi utilizzati per ottenere e rafforzare le immagini, come discusso in SEC06-BP02 Provision compute from hardened images.

È possibile utilizzare AWS Signer per la gestione della verifica delle firme, nonché del ciclo di vita di firma del codice per il tuo software e i tuoi artefatti. AWS Lambda e Amazon Elastic Container Registry offrono entrambi integrazioni con Signer per verificare le firme di codice e immagini. Utilizzando gli esempi nella sezione Risorse, puoi incorporare Signer nelle tue pipeline di integrazione e distribuzione continua (CI/CD) per automatizzare la verifica delle firme e la firma del tuo codice e delle tue immagini.

Risorse

Documenti correlati:

- Cryptographic Signing for Containers
- Le migliori pratiche per proteggere la pipeline di creazione delle immagini dei container utilizzando AWS Signer
- Annuncio della firma di Container Image con AWS Signer Amazon EKS
- Configurazione della firma del codice per AWS Lambda
- Best practices and advanced patterns for Lambda code signing
- Firma del codice tramite CA AWS Certificate Manager privata e chiavi AWS Key Management Service asimmetriche

Esempi correlati:

- Automatizza la firma del codice Lambda con Amazon e CodeCatalyst AWS Signer
- Firma e convalida OCI degli artefatti con AWS Signer

Strumenti correlati:

- AWS Lambda
- AWS Signer
- AWS Certificate Manager
- AWS Key Management Service
- AWS CodeArtifact

SEC06-BP05 Automatizza la protezione dell'elaborazione

Automatizza le operazioni di protezione delle risorse di calcolo per ridurre la necessità di intervento umano. Usa la scansione automatica per rilevare potenziali problemi all'interno delle tue risorse di calcolo e rimedia con risposte programmatiche automatiche o operazioni di gestione del parco. Incorpora l'automazione nei tuoi processi CI/CD per distribuire carichi di lavoro affidabili con dipendenze. up-to-date

Risultato desiderato: tutte le scansioni e le applicazioni di patch alle risorse di calcolo avvengono per mezzo di sistemi automatizzati. Utilizzate la verifica automatizzata per verificare che le immagini e le dipendenze del software provengano da fonti attendibili e non siano state manomesse. I carichi di lavoro vengono controllati automaticamente per individuare eventuali up-to-date dipendenze e firmati per stabilire l'affidabilità negli ambienti di elaborazione. AWS Le correzioni automatiche vengono avviate al rilevamento di risorse non conformi.

Anti-pattern comuni:

- Adozione della pratica dell'infrastruttura immutabile, senza però disporre di una soluzione di patch di emergenza o di sostituzione dei sistemi di produzione.
- Utilizzo dell'automazione per correggere le risorse non correttamente configurate, ma senza un meccanismo di annullamento manuale. Possono verificarsi situazioni in cui è necessario modificare i requisiti e sospendere le automazioni fino a quando non si modificano.

Vantaggi dell'adozione di questa best practice: riduzione del rischio di accessi alle risorse di calcolo e relativi utilizzi non autorizzati mediante l'automazione. Contribuisce a evitare che le configurazioni errate si diffondano negli ambienti di produzione e a rilevare e correggere tali configurazioni nel caso in cui si verifichino. L'automazione aiuta anche a rilevare l'accesso non autorizzato delle risorse di calcolo e il loro utilizzo, riducendo i tempi di risposta. In questo modo è possibile ridurre la portata complessiva dell'impatto del problema.

Livello di rischio associato se questa best practice non fosse adottata: medio

Guida all'implementazione

È possibile applicare le automazioni descritte nelle pratiche del pilastro della sicurezza per proteggere le risorse di calcolo. SEC06-BP01 Perform vulnerability management descrive come utilizzare Amazon Inspector nelle pipeline CI/CD e per la scansione continua degli ambienti di runtime alla ricerca di vulnerabilità ed esposizioni comuni note (). CVEs Puoi utilizzare AWS Systems Manager per applicare patch o eseguire nuove implementazioni da nuove immagini tramite runbook

automatizzati in modo da mantenere il tuo parco di calcolo aggiornato con software e librerie più recenti. Utilizza queste tecniche per ridurre la necessità di processi manuali e l'accesso interattivo alle tue risorse di elaborazione. SECPer saperne di più, consulta 06-BP03 Riduci la gestione manuale e l'accesso interattivo.

L'automazione svolge anche un ruolo nell'implementazione di carichi di lavoro affidabili, come descritto in SEC06-BP02 Fornire elaborazione da immagini rafforzate e 06-BP04 Convalidare l'integrità del software. SEC Puoi utilizzare servizi come EC2Image Builder e Amazon Elastic Container Registry (ECR) per scaricare, verificare, costruire e archiviare immagini e dipendenze di codice consolidate e approvate. AWS SignerAWS CodeArtifact Oltre a Inspector, ognuno di questi sistemi può svolgere un ruolo nel processo CI/CD, in modo che il carico di lavoro passi alla produzione solo quando viene confermato che le sue dipendenze provengono da fonti attendibili. up-to-date Il carico di lavoro è inoltre firmato in modo che gli ambienti di AWS calcolo, come AWS LambdaAmazon Elastic EKS Kubernetes Service (), possano verificare che non sia stato manomesso prima di consentirne l'esecuzione.

Oltre a questi controlli preventivi, è possibile utilizzare l'automazione nei controlli investigativi anche per le risorse di calcolo. Ad esempio, AWS Security Huboffre lo standard NIST800-53 Rev. 5 che include controlli come [EC2.8] EC2. Le istanze devono utilizzare Instance Metadata Service Version 2 (). IMDSv2 IMDSv2utilizza le tecniche di autenticazione della sessione, bloccando le richieste che contengono un' X-Forwarded-ForHTTPintestazione e una rete TTL di 1 per interrompere il traffico proveniente da fonti esterne e recuperare informazioni sull'istanza. EC2 Questo Security Hub di check-in può rilevare quando EC2 le istanze vengono utilizzate IMDSv1 e avviare la riparazione automatica. Scopri di più sul rilevamento e sulle riparazioni automatiche in SEC04-BP04 Avvia la correzione per le risorse non conformi.

Passaggi dell'implementazione

1. Automatizza la creazione sicura, conforme e avanzata con Image AMIs Builder. EC2 È possibile produrre immagini che incorporano i controlli degli standard dei benchmark del Center for Internet Security (CIS) o della Security Technical Implementation Guide (STIG) a partire da immagini di base e dei partner. AWS APN
2. Automatizza la gestione delle configurazioni. Applica e convalida in automatico le configurazioni sicure nelle risorse di calcolo utilizzando un servizio o uno strumento di gestione della configurazione.
   a. Gestione automatizzata della configurazione tramite AWS Config
   b. Gestione automatizzata del livello di sicurezza e conformità tramite AWS Security Hub

3. Automatizza l'applicazione di patch o la sostituzione delle istanze Amazon Elastic Compute Cloud (AmazonEC2). AWS Systems Manager Patch Manager automatizza il processo di applicazione di patch alle istanze gestite con aggiornamenti relativi alla sicurezza e di altro tipo. Gestione patch consente di applicare patch sia per i sistemi operativi sia per le applicazioni
   a. AWS Systems Manager Patch Manager
4. Automatizza la scansione delle risorse di elaborazione per individuare vulnerabilità ed esposizioni comuni (CVEs) e incorpora soluzioni di scansione di sicurezza all'interno della pipeline di sviluppo.
   a. Amazon Inspector
   b. ECRScansione di immagini
5. Prendi in considerazione Amazon GuardDuty per il rilevamento automatico di malware e minacce per proteggere le risorse di elaborazione. GuardDuty può anche identificare potenziali problemi quando una AWS Lambdafunzione viene richiamata nel tuo AWS ambiente.
   a. Amazon GuardDuty
6. Prendi in considerazione le soluzioni dei AWS partner. AWS I partner offrono prodotti leader del settore equivalenti, identici o integrati con i controlli esistenti negli ambienti locali. Questi prodotti integrano i servizi AWS esistenti per permettere di implementare un'architettura di sicurezza completa e un'esperienza più fluida nel cloud e negli ambienti on-premises.
   a. Sicurezza dell'infrastruttura

Risorse

Best practice correlate:

- SEC01-BP06 Automatizza l'implementazione dei controlli di sicurezza standard

Documenti correlati:

- Ottieni tutti i vantaggi e disabilita l'intera infrastruttura IMDSv2 IMDSv1 AWS

Video correlati:

- Best practice di sicurezza per il servizio di metadati delle EC2 istanze Amazon

Protezione dei dati

Questions

- SEC7. In che modo classifichi i dati?
- SEC8. Come proteggi i dati a riposo?
- SEC9. In che modo proteggi i dati in transito?

SEC7. In che modo classifichi i dati?

La classificazione fornisce un modo per categorizzare i dati in base ai livelli di criticità e sensibilità, in modo da aiutarti a determinare i controlli di protezione e conservazione appropriati.

Best practice

- SEC07-BP01 Comprendi il tuo schema di classificazione dei dati
- SEC07-BP02 Applica controlli di protezione dei dati basati sulla sensibilità dei dati
- SEC07-BP03 Identificazione e classificazione automatizzate
- SEC07-BP04 Definire una gestione scalabile del ciclo di vita dei dati

SEC07-BP01 Comprendi il tuo schema di classificazione dei dati

Comprendi la classificazione dei dati elaborati dal tuo carico di lavoro, i requisiti di gestione, i processi aziendali associati, dove sono archiviati i dati e chi è il relativo proprietario. Lo schema di classificazione e gestione dei dati deve tenere conto dei requisiti legali e di conformità applicabili del carico di lavoro e dei controlli dei dati necessari. Comprendere i dati è il primo passo nel percorso della classificazione dei dati.

Risultato desiderato: comprensione e documentazione ottimali dei tipi di dati presenti nel carico di lavoro. Sono in atto controlli adeguati per proteggere i dati sensibili in base alla loro classificazione. Questi controlli regolano considerazioni quali chi è autorizzato ad accedere ai dati e per quale scopo, la posizione di archiviazione dei dati, qual è la policy di crittografia per tali dati e le modalità di gestione delle chiavi di crittografia, il ciclo di vita dei dati e i requisiti di conservazione, i processi di distruzione opportuni, i processi di backup e ripristino in atto, nonché la verifica degli accessi.

Anti-pattern comuni:

- Non si dispone di una policy formale di classificazione dei dati per definire i livelli di sensibilità dei dati e i relativi requisiti di gestione.

- Non si dispone di una corretta consapevolezza dei livelli di sensibilità dei dati all'interno del carico di lavoro e non si acquisiscono queste informazioni nella documentazione dell'architettura e delle operazioni.
- Mancata applicazione di controlli appropriati sui dati in base alla loro sensibilità e ai requisiti, come indicato nella relativa policy di classificazione e trattamento.
- Mancata indicazione di un feedback sui requisiti di classificazione e trattamento dei dati ai proprietari delle policy.

Vantaggi dell'adozione di questa best practice: eliminazione delle ambiguità circa la corretta gestione dei dati nell'ambito del carico di lavoro grazie a questa pratica. L'applicazione di una policy formale che definisca i livelli di sensibilità dei dati nella propria organizzazione e le relative protezioni richieste, può aiutare a rispettare le normative legali e altre attestazioni e certificazioni di sicurezza informatica. I proprietari dei carichi di lavoro possono avere la certezza di sapere dove sono archiviati i dati sensibili e quali controlli di protezione sono in atto. La loro acquisizione nella documentazione aiuta i nuovi membri del team a comprenderli meglio e a gestire i controlli nelle prime fasi del loro mandato. Queste pratiche possono anche aiutare a ridurre i costi, dimensionando in modo corretto i controlli per ogni tipo di dati.

Livello di rischio associato se questa best practice non fosse adottata: elevato

Guida all'implementazione

Nella progettazione di un carico di lavoro, si può prendere in considerazione soluzioni per proteggere i dati sensibili in modo intuitivo. Ad esempio, in un'applicazione multi-tenant, è intuitivo considerare i dati di ciascun tenant come sensibili e mettere in atto protezioni in modo da vietare a un tenant l'accesso ai dati di un altro tenant. Allo stesso modo, è possibile progettare in modo intuitivo i controlli di accesso in modo che solo gli amministratori possano modificare i dati, e che gli altri utenti abbiano solo accesso a livello di lettura o non dispongano di alcun accesso.

Definizione e acquisizione di questi livelli di sensibilità dei dati nelle policy, insieme ai relativi requisiti di protezione dei dati, consente di identificare in modo formale la residenza dei dati nel tuo carico di lavoro. È quindi possibile determinare se sono stati predisposti i controlli giusti, se è possibile verificare i controlli e quali sono le risposte adeguate in caso di gestione errata dei dati.

Per agevolare la suddivisione in categorie delle aree con dati sensibili all'interno del carico di lavoro, valuta la possibilità di utilizzare i tag delle risorse, se disponibili. Ad esempio, puoi applicare un tag con una chiave di tag *`Classification`* e un valore di tag *`PHI`* per informazioni sanitarie protette (PHI) e un altro tag con una chiave di tag `Sensitivity` e un valore di tag di. `High` È possibile

usare servizi come AWS Config per monitorare tali risorse al fine di rilevare eventuali modifiche e inviare avvisi in caso di modifiche tali da renderle non conformi ai requisiti di protezione (come la modifica delle impostazioni di crittografia). È possibile acquisire la definizione standard delle chiavi tag e dei valori accettabili utilizzando le policy di tag, una funzionalità di AWS Organizations. Non è consigliabile che la chiave o il valore dei tag contenga dati privati o sensibili.

Passaggi dell'implementazione

1. Analizza lo schema di classificazione dei dati e i requisiti di protezione della tua organizzazione.
2. Identifica i tipi di dati sensibili elaborati dai tuoi carichi di lavoro.
3. Verifica che i dati sensibili siano archiviati e protetti all'interno del tuo carico di lavoro in base alla tua policy. Utilizza tecniche come i test automatizzati per verificare l'efficacia dei tuoi controlli.
4. Prendi in considerazione l'utilizzo di tag a livello di risorse e dati, laddove disponibili, per etichettare i dati con il relativo livello di sensibilità e altri metadati operativi che possono aiutare nel monitoraggio e nella risposta agli incidenti.
   a. AWS Organizations le politiche relative ai tag possono essere utilizzate per applicare gli standard di etichettatura.

Risorse

Best practice correlate:

- SUS04-BP01 Implementare una politica di classificazione dei dati

Documenti correlati:

- Whitepaper sulla classificazione dei dati
- Migliori pratiche per etichettare le risorse AWS

Esempi correlati:

- AWS Organizations Sintassi ed esempi della politica dei tag

Strumenti correlati

- Editor di tag AWS

SEC07-BP02 Applica controlli di protezione dei dati basati sulla sensibilità dei dati

Applica controlli di protezione dei dati che forniscano un livello di controllo adeguato a ciascuna classe di dati definita nella tua policy di classificazione. Questa pratica consente di proteggere i dati sensibili dall'accesso e dall'uso non autorizzati, preservandone al contempo disponibilità e utilizzo.

Risultato desiderato: presenza di una policy di classificazione che definisce i vari livelli di sensibilità dei dati nella tua organizzazione. Per ciascuno di questi livelli di sensibilità, disponi di linee guida chiare per servizi e luoghi di archiviazione e movimentazione approvati e per la loro configurazione richiesta. Implementi controlli per ciascun livello in base al livello di protezione richiesto e ai costi associati. Disponi di un sistema di monitoraggio e di avvisi per rilevare la presenza di dati in luoghi non autorizzati, l'elaborazione in ambienti non autorizzati, l'accesso da parte di soggetti non autorizzati o la configurazione di servizi correlati non conformi.

Anti-pattern comuni:

- Applicazione dello stesso livello di controlli di protezione su tutti i dati. Ciò può portare a un eccesso di controlli di sicurezza per i dati a bassa sensibilità o a una protezione insufficiente dei dati altamente sensibili.
- Mancato coinvolgimento delle parti interessate dei team di sicurezza, conformità e business nella definizione dei controlli sulla protezione dei dati.
- Si trascurano le spese generali e i costi operativi associati all'implementazione e al mantenimento dei controlli sulla protezione dei dati.
- Mancata effettuazione di revisioni periodiche del controllo della protezione dei dati per mantenere l'allineamento con le policy di classificazione.

Vantaggi dell'adozione di questa best practice: grazie all'allineamento dei controlli al livello di classificazione dei dati, l'organizzazione può investire in livelli di controllo più elevati, laddove necessario. Ciò può includere l'aumento delle risorse per la sicurezza, il monitoraggio, la misurazione, la correzione e la creazione di report. Se è opportuno disporre di meno controlli, è possibile migliorare l'accessibilità e la completezza dei dati per il personale, i clienti o gli utenti. Questo approccio offre alla tua organizzazione la massima flessibilità nell'utilizzo dei dati, pur rispettandone i requisiti di protezione.

Livello di rischio associato se questa best practice non fosse adottata: elevato

Guida all'implementazione

L'implementazione dei controlli di protezione dei dati in base ai loro livelli di sensibilità comporta diverse fasi fondamentali. In promo luogo, identifica i diversi livelli di sensibilità dei dati all'interno dell'architettura del tuo carico di lavoro (ad esempio, pubblico, interno, riservato e limitato) e valuta il luogo in cui memorizzi ed elabori questi dati. Quindi, definisci i limiti di isolamento dei dati in base al loro livello di sensibilità. Ti consigliamo di separare i dati in diversi Account AWS, utilizzando le policy di controllo dei servizi (SCPs) per limitare i servizi e le azioni consentiti per ogni livello di sensibilità dei dati. In questo modo, puoi creare forti limiti di isolamento e far rispettare il principio del privilegio minimo.

Una volta definiti i limiti di isolamento, implementa i controlli di protezione adeguati in base ai loro livelli di sensibilità. Consulta le best practice per la protezione dei dati a riposo e la protezione dei dati in transito in modo da implementare controlli pertinenti come la crittografia, i controlli di accesso e gli audit. Prendi in considerazione tecniche come la tokenizzazione o l'anonimizzazione per ridurre il livello di sensibilità dei tuoi dati. Semplifica l'applicazione di policy coerenti sui dati in tutta l'azienda con un sistema centralizzato per la tokenizzazione e la de-tokenizzazione.

Monitora e verifica in modo continuo l'efficacia dei controlli implementati. Rivedi e aggiorna con regolarità lo schema di classificazione dei dati, le valutazioni dei rischi e i controlli di protezione in base all'evoluzione del panorama di dati e minacce dell'organizzazione. Allinea i controlli di protezione dei dati implementati con normative, standard e requisiti legali pertinenti del settore. Inoltre, procedi alla sensibilizzazione e formazione sulla sicurezza per aiutare i dipendenti a comprendere lo schema di classificazione dei dati e le loro responsabilità nella gestione e protezione dei dati sensibili.

Passaggi dell'implementazione

1. Identifica i livelli di classificazione e sensibilità dei dati all'interno del tuo carico di lavoro.
2. Definisci i limiti di isolamento per ciascun livello e determina una strategia di applicazione.
3. Valuta i controlli definiti che regolano accesso, crittografia, verifica, conservazione e altri aspetti richiesti dalla policy di classificazione dei dati.
4. Valuta le opzioni per ridurre il livello di sensibilità dei dati laddove appropriato, ad esempio utilizzando la tokenizzazione o l'anonimizzazione.
5. Verifica i tuoi controlli utilizzando test e monitoraggio automatici delle risorse configurate.

Risorse

Best practice correlate:

- PERF03-BP01 Utilizza un archivio dati appositamente progettato che supporti al meglio i requisiti di accesso e archiviazione dei dati
- COST04-BP05 Applica le politiche di conservazione dei dati

Documenti correlati:

- Whitepaper sulla classificazione dei dati
- Best practice per la sicurezza, l'identità e la conformità
- AWS KMS Best practice
- Le migliori pratiche e funzionalità di crittografia per i servizi AWS

Esempi correlati:

- Building a serverless tokenization solution to mask sensitive data
- How to use tokenization to improve data security and reduce audit scope

Strumenti correlati:

- AWS Key Management Service (AWS KMS)
- AWS CloudHSM
- AWS Organizations

SEC07-BP03 Identificazione e classificazione automatizzate

Automatizzare l'identificazione e la classificazione dei dati può aiutarti a implementare i controlli corretti. L'uso dell'automazione per aumentare la determinazione manuale riduce il rischio di errore umano e di esposizione.

Risultato desiderato: possibilità di verificare se sono in atto controlli adeguati in base alla policy di classificazione e gestione. Strumenti e servizi automatizzati ti aiutano a identificare e classificare il livello di sensibilità dei tuoi dati. L'automazione consente inoltre di monitorare in modo continuo

gli ambienti in modo da rilevare e inviare avvisi se i dati vengono archiviati o gestiti in modo non autorizzato, così da poter intraprendere rapidamente azioni correttive.

Anti-pattern comuni:

- Affidarsi esclusivamente a processi manuali per l'identificazione e la classificazione dei dati, che possono essere soggetti a errori e richiedere tempi di lavoro lunghi. Questo può portare a una classificazione dei dati inefficiente e incoerente, soprattutto con l'aumento dei volumi di dati.
- Mancata predisposizione di meccanismi per tracciare e gestire le risorse di dati all'interno dell'organizzazione.
- Si trascura la necessità di un monitoraggio e di una classificazione continui dei dati durante i loro spostamenti e le loro trasformazioni all'interno dell'organizzazione.

Vantaggi dell'adozione di questa best practice: l'automazione di identificazione e classificazione dei dati può garantire un'applicazione più coerente e accurata dei controlli di protezione dei dati, così da ridurre il rischio di errore umano. L'automazione può inoltre fornire visibilità in merito ad accesso e movimento dei dati sensibili, così da rilevare le manipolazioni non autorizzate e intraprendere azioni correttive.

Livello di rischio associato se questa best practice non fosse adottata: medio

Guida all'implementazione

Sebbene si ricorra spesso al giudizio umano per classificare i dati durante le fasi iniziali di progettazione di un carico di lavoro, è opportuno considerare la presenza di sistemi che automatizzino l'identificazione e la classificazione dei dati di test come controllo preventivo. Ad esempio, agli sviluppatori può essere fornito uno strumento o un servizio per analizzare i dati rappresentativi e determinarne la sensibilità. All'interno AWS, puoi caricare set di dati in Amazon S3 e scansionarli utilizzando Amazon Macie, Amazon Comprehend o Amazon Comprehend Medical. Allo stesso modo, considera la scansione dei dati come parte dei test di unità e integrazione per individuare i casi in cui i dati sensibili non sono previsti. Gli avvisi sui dati sensibili in questa fase possono evidenziare le lacune nelle protezioni prima dell'implementazione in produzione. Altre funzionalità, come il rilevamento di dati sensibili in AWS GlueSNS, Amazon e Amazon, CloudWatch possono essere utilizzate anche per rilevare PII e adottare misure di mitigazione. Per qualsiasi strumento o servizio automatizzato, esamina come definisce i dati sensibili e integralo con altre soluzioni umane o automatizzate per colmare eventuali lacune.

Come controllo investigativo, utilizza il monitoraggio continuo degli ambienti per rilevare l'eventuale archiviazione non conforme dei dati sensibili. In questo modo puoi rilevare situazioni come l'emissione di dati sensibili nei file di log o la loro copia in un ambiente di analisi dei dati senza un'adeguata de-identificazione o redazione. I dati archiviati in Amazon S3 possono essere costantemente monitorati per verificare la presenza di dati sensibili grazie ad Amazon Macie.

Passaggi dell'implementazione

1. Esegui una scansione iniziale degli ambienti per l'identificazione e la classificazione automatica.
    a. Una prima scansione completa dei dati può aiutare a capire la residenza dei dati sensibili nei tuoi ambienti. Qualora una scansione completa non sia inizialmente richiesta o non possa essere completata in anticipo a causa dei costi, valuta l'adeguatezza delle tecniche di campionamento per raggiungere i tuoi risultati. Ad esempio, Amazon Macie può essere configurato per eseguire un'ampia operazione automatizzata di rilevamento dei dati sensibili nei bucket S3. Questa funzionalità utilizza tecniche di campionamento per eseguire in modo efficiente in termini di costi un'analisi preliminare della residenza dei dati. È quindi possibile eseguire un'analisi più approfondita dei bucket S3 utilizzando un processo di rilevamento dei dati sensibili. Anche altri archivi di dati possono essere esportati su S3 per essere analizzati da Macie.
2. Configura scansioni continue dei tuoi ambienti.
    a. La capacità di rilevamento automatizzata dei dati sensibili di Macie consente di eseguire scansioni continue degli ambienti. I bucket S3 noti e autorizzati a memorizzare dati sensibili possono essere esclusi utilizzando un elenco di permessi in Macie.
3. Incorpora l'identificazione e la classificazione nei processi di compilazione e di test.
    a. Identifica gli strumenti utilizzabili dagli sviluppatori per analizzare i dati alla ricerca di sensibilità mentre i carichi di lavoro sono in fase di sviluppo. Utilizza questi strumenti come parte dei test di integrazione per avvisare quando i dati sensibili sono inaspettati e impedire un'ulteriore implementazione.
4. Implementa un sistema o un runbook per intervenire quando i dati sensibili vengono trovati in luoghi non autorizzati.

Risorse

Documenti correlati:

- AWS Glue: Detect and process sensitive data

- Utilizzo di identificatori di dati gestiti in Amazon SNS
- Amazon CloudWatch Logs: aiuta a proteggere i dati di log sensibili con il mascheramento

Esempi correlati:

- Abilitazione della classificazione dei dati per il RDS database Amazon con Macie
- Detecting sensitive data in DynamoDB with Macie

Strumenti correlati:

- Amazon Macie
- Amazon Comprehend
- Amazon Comprehend Medical
- AWS Glue

SEC07-BP04 Definire una gestione scalabile del ciclo di vita dei dati

Esamina i requisiti del ciclo di vita dei dati in relazione ai loro diversi livelli di classificazione e gestione. Ciò può includere le modalità di gestione dei dati quando entrano per la prima volta nell'ambiente, il modo in cui i dati si trasformano e le regole per la loro distruzione. Prendi in considerazione fattori come periodi di conservazione, accesso, audit e monitoraggio della provenienza.

Risultato desiderato: classificazione dei dati il più vicino possibile al momento e all'ora dell'importazione. Quando la classificazione dei dati richiede il mascheramento, la tokenizzazione o altri processi che riducono il livello di sensibilità, si eseguono queste azioni il più vicino possibile al punto e al momento dell'importazione.

Elimini i dati in conformità con la policy in uso quando non è più opportuno conservarli, in base alla loro classificazione.

Anti-pattern comuni:

- Implementazione di un one-size-fits-all approccio alla gestione del ciclo di vita dei dati, senza considerare i diversi livelli di sensibilità e i requisiti di accesso.
- Valutazione della gestione del ciclo di vita solo dal punto di vista dei dati utilizzabili o dei dati di cui si esegue il backup, ma non di entrambi.

- Si presume che i dati immessi nel carico di lavoro siano validi, senza stabilirne il valore o la provenienza.
- Affidamento alla durabilità dei dati come sostituti dei backup e della protezione dei dati.
- Mantenimento dei dati oltre la loro utilità e il periodo di conservazione richiesto.

Vantaggi dell'adozione di questa best practice: una strategia di gestione del ciclo di vita dei dati ben definita e scalabile aiuta a mantenere la conformità normativa, migliora la sicurezza dei dati, ottimizza i costi di archiviazione e consente l'accesso e la condivisione efficienti dei dati mantenendo i controlli opportuni.

Livello di rischio associato se questa best practice non fosse adottata: elevato

Guida all'implementazione

I dati all'interno di un carico di lavoro sono spesso dinamici. La forma che assumono quando entrano nell'ambiente del carico di lavoro può essere diversa da quella che assumono quando vengono archiviati o utilizzati nella logica aziendale, nel reporting, nell'analisi o nel machine learning. Inoltre, il valore dei dati può cambiare nel tempo. Alcuni dati sono di natura temporale e perdono valore con il passare del tempo. Considera l'impatto di queste modifiche ai dati sulla valutazione del tuo schema di classificazione dei dati e dei controlli associati. Laddove possibile, utilizza un meccanismo automatizzato del ciclo di vita, come le policy del ciclo di vita di Amazon S3 e Amazon Data Lifecycle Manager, per configurare i processi di scadenza, archiviazione e conservazione dei dati.

Distingui tra i dati disponibili per l'uso e quelli archiviati come backup. Prendi in considerazione l'utilizzo AWS Backupper automatizzare il backup dei dati tra i servizi. AWS EBSLe istantanee di Amazon forniscono un modo per copiare un EBS volume e archiviarlo utilizzando le funzionalità di S3, tra cui ciclo di vita, protezione dei dati e accesso ai meccanismi di protezione. Due di questi meccanismi sono S3 Object Lock e AWS Backup Vault Lock, in grado di garantire sicurezza e controllo aggiuntivi ai backup. Gestisci una chiara separazione dei compiti e dell'accesso per i backup. Isola i backup a livello di account per mantenere la separazione dall'ambiente interessato durante un evento.

Un altro aspetto della gestione del ciclo di vita consiste nella registrazione della cronologia dei dati mentre avanzano nel carico di lavoro, chiamato tracciamento della provenienza dei dati. In questo modo hai la certezza di conoscere la provenienza dei dati, le trasformazioni effettuate, il proprietario o il processo che ha apportato le modifiche e la data. Questa cronologia è utile per la risoluzione dei problemi e le analisi in caso di potenziali eventi di sicurezza. Ad esempio, puoi creare log sui

metadati relativi alle trasformazioni in una tabella Amazon DynamoDB. All'interno di un data lake, puoi conservare copie dei dati trasformati in diversi bucket S3 per ciascuna fase della pipeline di dati. Archivia le informazioni su schema e timestamp in un AWS Glue Data Catalog. Indipendentemente dalla tua soluzione, considera i requisiti degli utenti finali per determinare gli strumenti appropriati di cui hai bisogno per segnalare la provenienza dei tuoi dati. In questo modo potrai determinare come tracciare al meglio la tua provenienza.

Passaggi dell'implementazione

1. Analizza i tipi di dati, i livelli di sensibilità e i requisiti di accesso del carico di lavoro per classificare i dati e definire strategie di gestione del ciclo di vita appropriate.
2. Progetta e implementa policy di conservazione dei dati e processi di distruzione automatizzata in linea con i requisiti legali, normativi e organizzativi.
3. Stabilisci processi e automazione per il monitoraggio continuo, la verifica e l'adeguamento delle strategie, dei controlli e delle policy di gestione del ciclo di vita dei dati in base all'evoluzione dei requisiti del carico di lavoro e delle normative.

Risorse

Best practice correlate:

- COST04-BP05 Applica le politiche di conservazione dei dati
- SUS04-BP03 Utilizza le policy per gestire il ciclo di vita dei tuoi set di dati

Documenti correlati:

- Whitepaper sulla classificazione dei dati
- AWS Blueprint for Ransomware Defense
- DevOpsGuida: migliora la tracciabilità con il monitoraggio della provenienza dei dati

Esempi correlati:

- How to protect sensitive data for its entire lifecycle in AWS
- Crea un data lineage per i data lake utilizzando AWS Glue Amazon Neptune e Spline

Strumenti correlati:

- AWS Backup
- Amazon Data Lifecycle Manager
- AWS Identity and Access Management Access Analyzer

SEC8. Come proteggi i dati a riposo?

Proteggi i dati a riposo implementando più controlli per ridurre il rischio di accessi non autorizzati o altri comportamenti impropri.

Best practice

- SEC08-BP01 Implementare una gestione sicura delle chiavi
- SEC08-BP02 Applica la crittografia a riposo
- SEC08-BP03 Automatizza la protezione dei dati a riposo
- SEC08-BP04 Applica il controllo degli accessi

SEC08-BP01 Implementare una gestione sicura delle chiavi

La gestione sicura delle chiavi include l'archiviazione, la rotazione, il controllo degli accessi e il monitoraggio del materiale relativo alla chiave necessario per proteggere i dati a riposo per il carico di lavoro.

Risultato desiderato: un meccanismo di gestione delle chiavi scalabile, ripetibile e automatizzato. Il meccanismo dovrebbe fornire la possibilità di applicare l'accesso con il privilegio minimo al materiale relativo alla chiave e offrire il giusto equilibrio tra disponibilità, riservatezza e integrità delle chiavi. L'accesso alle chiavi va monitorato e occorre ruotare il materiale relativo alla chiave mediante un processo automatizzato. Il materiale relativo alla chiave non dovrebbe mai essere accessibile alle identità umane.

Anti-pattern comuni:

- Accesso umano a materiale relativo alla chiave non crittografato.
- Creazione di algoritmi crittografici personalizzati.
- Autorizzazioni di accesso al materiale relativo alla chiave di accesso troppo ampie.

Vantaggi dell'adozione di questa best practice: predisponendo un meccanismo di gestione delle chiavi sicuro per il tuo carico di lavoro, puoi contribuire a proteggere i contenuti dagli accessi non

autorizzati. Inoltre, la crittografia dei dati potrebbe essere prevista da requisiti normativi per la tua organizzazione. Una soluzione efficace di gestione delle chiavi può fornire meccanismi tecnici finalizzati alla protezione del materiale relativo alle chiavi in linea con tali normative.

Livello di rischio associato se questa best practice non fosse adottata: elevato

Guida all'implementazione

Molti requisiti normativi e best practice includono la crittografia dei dati a riposo come controllo di sicurezza fondamentale. Per garantire la conformità a questo controllo, il carico di lavoro necessita di un meccanismo per archiviare e gestire in modo sicuro il materiale relativo alla chiave utilizzato per crittografare i dati a riposo.

AWS offre AWS Key Management Service (AWS KMS) per fornire uno spazio di archiviazione durevole, sicuro e ridondante per le chiavi. AWS KMS Molti AWS servizi si integrano con AWS KMS per supportare la crittografia dei dati. AWS KMS utilizza moduli di sicurezza hardware convalidati FIPS 140-2 di livello 3 per proteggere le chiavi. Non esiste alcun meccanismo per esportare AWS KMS le chiavi in testo semplice.

Quando si distribuiscono carichi di lavoro utilizzando una strategia multi-account, è consigliabile conservare AWS KMS le chiavi nello stesso account del carico di lavoro che le utilizza. In questo modello distribuito, la responsabilità della gestione delle AWS KMS chiavi spetta al team dell'applicazione. In altri casi d'uso, le organizzazioni possono scegliere di archiviare AWS KMS le chiavi in un account centralizzato. Questa struttura centralizzata richiede policy aggiuntive per consentire l'accesso multi-account richiesto affinché l'account del carico di lavoro possa accedere alle chiavi di accesso archiviate nell'account centralizzato, ma può essere più applicabile nei casi d'uso in cui una singola chiave è condivisa tra Account AWS multipli.

Indipendentemente dal luogo in cui è archiviato il materiale chiave, l'accesso alla chiave deve essere strettamente controllato mediante l'uso di politiche e IAM politiche chiave. Le politiche chiave sono il modo principale per controllare l'accesso a una AWS KMS chiave. Inoltre, le concessioni di AWS KMS chiavi possono fornire l'accesso ai AWS servizi per crittografare e decrittografare i dati per conto dell'utente. Prenditi del tempo per esaminare le migliori pratiche per il controllo degli accessi alle tue chiavi. AWS KMS

Una best practice è quella di monitorare l'uso delle chiavi di crittografia per rilevare modelli di accesso insoliti. Le operazioni eseguite utilizzando chiavi AWS gestite e chiavi gestite dal cliente archiviate in AWS KMS possono essere registrate AWS CloudTrail e devono essere riviste periodicamente. Occorre prestare particolare attenzione al monitoraggio dei principali eventi di eliminazione delle

chiavi. Per ridurre le probabilità di distruzione accidentale o dolosa del materiale relativo alla chiave, gli eventi di eliminazione delle chiavi non hanno efficacia immediata. I tentativi di eliminare le chiavi in AWS KMS ingresso sono soggetti a un periodo di attesa, che per impostazione predefinita è di 30 giorni, che offre agli amministratori il tempo di esaminare queste azioni e annullare la richiesta, se necessario.

La maggior parte dei AWS servizi utilizza AWS KMS in modo trasparente per l'utente: l'unico requisito è decidere se utilizzare una chiave AWS gestita o gestita dal cliente. Se il carico di lavoro richiede l'uso diretto di AWS KMS per crittografare o decrittografare i dati, la migliore pratica è utilizzare la crittografia a busta per proteggere i dati. The AWS Encryption SDK può fornire alle applicazioni primitive di crittografia lato client con cui implementare la crittografia in busta e con cui integrarsi. AWS KMS

Passaggi dell'implementazione

1. Determinate le opzioni di gestione delle chiavi appropriate (AWS gestite o gestite dal cliente) per la chiave.
   - Per facilitare l'uso, AWS offre chiavi AWS possedute e AWS gestite per la maggior parte dei servizi, che forniscono encryption-at-rest funzionalità senza la necessità di gestire il materiale o le politiche chiave.
   - Quando utilizzi chiavi gestite dal cliente, prendi in considerazione il keystore predefinito per fornire il miglior equilibrio tra agilità, sicurezza, sovranità dei dati e disponibilità. Per altri casi d'uso può essere richiesto l'uso di archivi di chiavi personalizzati con AWS CloudHSM o l'archivio chiavi esterno.
2. Consulta l'elenco dei servizi che stai utilizzando per il tuo carico di lavoro per capire come AWS KMS si integra con il servizio. Ad esempio, EC2 le istanze possono utilizzare EBS volumi crittografati, verificando che anche le EBS istantanee Amazon create da tali volumi siano crittografate utilizzando una chiave gestita dal cliente e mitigando la divulgazione accidentale di dati di snapshot non crittografati.
   - Come vengono utilizzati i servizi AWSAWS KMS
   - Per informazioni dettagliate sulle opzioni di crittografia offerte da un AWS servizio, consulta l'argomento Encryption at Rest nella guida per l'utente o la guida per sviluppatori del servizio.
3. Implementazione AWS KMS: AWS KMS semplifica la creazione e la gestione delle chiavi e il controllo dell'uso della crittografia in un'ampia gamma di AWS servizi e nelle applicazioni.
   - Guida introduttiva: AWS Key Management Service (AWS KMS)
   - Esamina le migliori pratiche per il controllo degli accessi alle tue AWS KMS chiavi.

4. Considera AWS Encryption SDK: usa AWS Encryption SDK with AWS KMS integration quando la tua applicazione deve crittografare i dati lato client.
   - AWS Encryption SDK
5. Abilita IAMAccess Analyzer per esaminare e notificare automaticamente se esistono politiche chiave troppo ampie. AWS KMS
6. Abilita Security Hub per ricevere notifiche in caso di policy della chiave configurate in modo errato, chiavi programmate per essere eliminate o chiavi senza la rotazione automatica abilitata.
7. Determina il livello di registrazione appropriato per le tue chiavi. AWS KMS Poiché le chiamate a AWS KMS, compresi gli eventi di sola lettura, vengono registrate, i CloudTrail registri associati possono diventare voluminosi. AWS KMS
   - Alcune organizzazioni preferiscono separare l'attività di registrazione in un percorso separato. AWS KMS Per maggiori dettagli, consulta la CloudTrail sezione Registrazione delle AWS KMS API chiamate con della guida per gli sviluppatori. AWS KMS

Risorse

Documenti correlati:

- AWS Key Management Service
- AWS cryptographic services and tools
- Protezione dei dati Amazon S3 tramite la crittografia
- Crittografia envelope
- Impegno per la sovranità digitale
- Demystifying AWS KMS key operations, bring your own key, custom key store, and ciphertext portability
- AWS Key Management Service dettagli crittografici

Video correlati:

- Come funziona la crittografia in AWS
- Protezione dello storage a blocchi su AWS
- AWS data protection: Using locks, keys, signatures, and certificates

Esempi correlati:

- Implementa meccanismi avanzati di controllo degli accessi utilizzando AWS KMS

SEC08-BP02 Applica la crittografia a riposo

Per i dati a riposo è necessario applicare la crittografia. La crittografia mantiene la riservatezza dei dati sensibili in caso di accesso non autorizzato o di divulgazione accidentale.

Risultato desiderato: per impostazione predefinita, si applica la crittografia ai dati privati quando sono a riposo. La crittografia aiuta a mantenere la riservatezza dei dati e fornisce un ulteriore livello di protezione contro la divulgazione o esfiltrazione intenzionale o involontaria dei dati. I dati crittografati non possono essere letti o consultati senza che siano stati prima decrittografati. Tutti i dati archiviati in modo non crittografato devono essere inventariati e controllati.

Anti-pattern comuni:

- Non encrypt-by-default utilizza configurazioni.
- Accesso estremamente permissivo alle chiavi di decrittografia.
- Mancato monitoraggio dell'uso delle chiavi di crittografia e decrittografia.
- Memorizzazione di dati non crittografati.
- Utilizzo della stessa chiave di crittografia per tutti i dati, indipendentemente dall'uso, dal tipo e dalla classificazione dei dati stessi.

Livello di rischio associato se questa best practice non fosse adottata: elevato

Guida all'implementazione

Mappa le chiavi di crittografia in base alle classificazioni dei dati all'interno dei carichi di lavoro. Questo approccio favorisce la protezione dei dati da accessi eccessivamente permissivi in caso di utilizzo di una sola chiave di crittografia o di un numero molto ridotto di chiavi di crittografia (vedi SEC07-BP01 Comprendi il tuo schema di classificazione dei dati).

AWS Key Management Service (AWS KMS) si integra con molti AWS servizi per semplificare la crittografia dei dati inattivi. Ad esempio, in Amazon Simple Storage Service (Amazon S3) puoi impostare la crittografia predefinita su un bucket in modo che tutti nuovi oggetti vengano crittografati in automatico. Durante l'utilizzo AWS KMS, considera quanto strettamente i dati devono essere limitati. Le AWS KMS chiavi predefinite e controllate dal servizio vengono gestite e utilizzate per tuo conto da. AWS Per i dati sensibili che richiedono un accesso granulare alla chiave di crittografia

sottostante, prendi in considerazione le chiavi gestite dal cliente (). CMKs Hai il pieno controlloCMKs, inclusa la rotazione e la gestione degli accessi, tramite l'uso di politiche chiave.

Inoltre, Amazon Elastic Compute Cloud (AmazonEC2) e Amazon S3 supportano l'applicazione della crittografia impostando la crittografia predefinita. È possibile utilizzarla Regole di AWS Configper verificare automaticamente che si stia utilizzando la crittografia, ad esempio per volumi Amazon Elastic Block Store (AmazonEBS), istanze Amazon Relational Database Service (RDSAmazon) e bucket Amazon S3.

AWS fornisce anche opzioni per la crittografia lato client, che consentono di crittografare i dati prima di caricarli sul cloud. AWS Encryption SDK Fornisce un modo per crittografare i dati utilizzando la crittografia a busta. Fornisci la chiave di wrapping e poi AWS Encryption SDK genera una chiave dati unica per ogni oggetto di dati che crittografa. Valuta AWS CloudHSM se hai bisogno di un modulo di sicurezza hardware a tenant singolo gestito (). HSM AWS CloudHSM consente di generare, importare e gestire chiavi crittografiche su un sistema FIPS 140-2 di livello 3 convalidato. HSM Alcuni casi d'uso AWS CloudHSM includono la protezione delle chiavi private per il rilascio di un'autorità di certificazione (CA) e l'attivazione della crittografia trasparente dei dati (TDE) per i database Oracle. Il AWS CloudHSM client SDK fornisce un software che consente di crittografare i dati lato client utilizzando le chiavi memorizzate all'interno AWS CloudHSM prima del caricamento dei dati. AWS La crittografia lato client Amazon DynamoDB consente inoltre di crittografare e firmare gli elementi prima del caricamento in una tabella DynamoDB.

Passaggi dell'implementazione

- Applica la crittografia a riposo per Amazon S3: implementa la crittografia predefinita del bucket Amazon S3..

  Configura la crittografia predefinita per i nuovi EBS volumi Amazon: specifica che desideri che tutti i EBS volumi Amazon appena creati vengano creati in forma crittografata, con la possibilità di utilizzare la chiave predefinita fornita da AWS o una chiave creata da te.

  Configurazione di Amazon Machine Images crittografate (AMIs): la copia di un file esistente AMI con crittografia configurata crittograferà automaticamente i volumi root e gli snapshot.

  Configura RDSla crittografia Amazon: configura la crittografia per i cluster di RDS database Amazon e gli snapshot inattivi utilizzando l'opzione di crittografia.

  Crea e configura AWS KMS chiavi con politiche che limitano l'accesso ai principi appropriati per ogni classificazione di dati: ad esempio, crea una AWS KMS chiave per crittografare i dati di produzione e una chiave diversa per crittografare i dati di sviluppo o di test. Puoi anche

fornire l'accesso tramite chiave ad altri. Account AWS Considera la possibilità di predisporre account diversi per gli ambienti di sviluppo e di produzione. Se l'ambiente di produzione deve decrittografare gli artefatti nell'account di sviluppo, puoi modificare la CMK politica utilizzata per crittografare gli artefatti di sviluppo per consentire all'account di produzione di decrittografare tali artefatti. L'ambiente di produzione può quindi importare i dati decrittografati per utilizzarli nella produzione.

Configura la crittografia in AWS servizi aggiuntivi: per gli altri AWS servizi che utilizzi, consulta la documentazione sulla sicurezza relativa a quel servizio per determinare le opzioni di crittografia del servizio.

Risorse

Documenti correlati:

- AWS Crypto Tools
- AWS Encryption SDK
- AWS KMS Whitepaper sui dettagli crittografici
- AWS Key Management Service
- AWS cryptographic services and tools
- EBSCrittografia Amazon
- Crittografia predefinita per EBS i volumi Amazon
- Crittografia delle risorse Amazon RDS
- Come si attiva la crittografia predefinita per un bucket Amazon S3?
- Protezione dei dati Amazon S3 tramite la crittografia

Video correlati:

- Come funziona la crittografia in AWS
- Protezione dello storage a blocchi su AWS

SEC08-BP03 Automatizza la protezione dei dati a riposo

Usa l'automazione per convalidare e applicare i controlli dei dati a riposo. Usa la scansione automatica per rilevare le configurazioni errate delle soluzioni di archiviazione di dati ed esegui le

correzioni attraverso la risposta programmatica automatica, ove possibile. Incorpora l'automazione nei tuoi processi CI/CD per rilevare le configurazioni errate dell'archiviazione di dati prima che vengano implementate in produzione.

Risultato desiderato: scansione e monitoraggio da parte di sistemi automatizzati delle posizioni di archiviazione di dati per individuare configurazioni errate dei controlli, accessi non autorizzati e usi imprevisti. Il rilevamento delle posizioni di archiviazione non configurate avvia correzioni automatiche. I processi automatizzati creano backup dei dati e archiviano copie immutabili al di fuori dell'ambiente originale.

Anti-pattern comuni:

- Mancata tenuta in considerazione delle opzioni per abilitare la crittografia dalle impostazioni predefinite, ove supportate.
- Mancata tenuta in considerazione degli eventi di sicurezza, oltre a quelli operativi, quando si formula una strategia di backup e ripristino automatizzata.
- Mancata applicazione delle impostazioni di accesso pubblico per i servizi di archiviazione.
- Assenza di monitoraggio e audit dei controlli per proteggere i dati a riposo.

Vantaggi dell'adozione di questa best practice: prevenzione grazie all'automazione del rischio di configurazioni errate delle posizioni di archiviazione di dati e dell'ingresso di configurazioni errate negli ambienti di produzione. Questa best practice aiuta anche a rilevare e correggere eventuali configurazioni errate.

Livello di rischio associato se questa best practice non fosse adottata: medio

Guida all'implementazione

L'automazione è un tema ricorrente in tutte le pratiche per la protezione dei dati a riposo. SEC01-BP06 L'implementazione automatica dei controlli di sicurezza standard descrive come acquisire la configurazione delle risorse utilizzando modelli di infrastruttura come codice (IaC), ad esempio con. AWS CloudFormation Questi modelli sono affidati a un sistema di controllo della versione e vengono utilizzati per distribuire risorse tramite una pipeline CI/CD. AWS Queste tecniche si applicano anche all'automazione della configurazione delle soluzioni di archiviazione di dati, come le impostazioni di crittografia sui bucket Amazon S3.

Puoi controllare le impostazioni che definisci nei tuoi modelli IaC per eventuali configurazioni errate nelle pipeline CI/CD utilizzando le regole in AWS CloudFormation Guard. È possibile monitorare

impostazioni che non sono ancora disponibili in altri strumenti IAc per rilevare CloudFormation eventuali errori di configurazione con. AWS Config Gli avvisi generati da Config per configurazioni errate possono essere corretti automaticamente, come descritto SECin 04-BP04 Avviare la riparazione per risorse non conformi.

L'utilizzo dell'automazione come parte della strategia di gestione delle autorizzazioni è anche parte integrante delle protezioni automatizzate dei dati. SEC03-BP02 Garantisci l'accesso con il minimo privilegio e SEC03-BP04 Reduce le autorizzazioni descrivono continuamente la configurazione delle politiche di accesso con privilegi minimi che vengono continuamente monitorate da la per generare risultati quando l'autorizzazione può essere ridotta. AWS Identity and Access Management Access Analyzer Oltre GuardDuty all'automazione per il monitoraggio delle autorizzazioni, puoi configurare Amazon per rilevare comportamenti anomali di accesso ai dati per i tuoi EBSvolumi (a titolo di EC2 istanza), i bucket S3 e i database Amazon Relational Database Service supportati.

L'automazione svolge rileva inoltre i casi di archiviazione di dati sensibili in luoghi non autorizzati. SEC07-BP03 Identificazione e classificazione automatizzate descrive in che modo Amazon Macie può monitorare i bucket S3 alla ricerca di dati sensibili imprevisti e generare avvisi in grado di avviare una risposta automatica.

Segui le pratiche riportate in REL09 Esegui il backup dei dati per sviluppare una strategia automatizzata di backup e ripristino dei dati. Il backup e il ripristino dei dati sono importanti tanto per il ripristino da eventi di sicurezza quanto per gli eventi operativi.

Passaggi dell'implementazione

1. Acquisisci la configurazione dell'archiviazione di dati nei modelli IaC. Utilizza i controlli automatizzati nelle pipeline CI/CD per rilevare configurazioni errate.
   a. Puoi usare per i tuoi modelli IaC e CloudFormationGuard per verificare eventuali errori di configurazione dei modelli.
   b. Utilizza AWS Config per eseguire le regole in modalità di valutazione proattiva. Utilizza questa impostazione per verificare la conformità di una risorsa come passaggio della pipeline CI/CD prima di crearla.
2. Monitora le risorse per individuare eventuali configurazioni errate dell'archiviazione di dati.
   a. Imposta AWS Config in modo che monitori le risorse di archiviazione di dati al fine di rilevare eventuali modifiche nelle configurazioni di controllo e generare avvisi per richiamare correzioni in caso di rilevamento di una configurazione errata.
   b. Vedi SEC04-BP04 Avviare la riparazione per risorse non conformi per ulteriori indicazioni sulle riparazioni automatiche.

3. Monitora e riduci in modo continuo le autorizzazioni di accesso ai dati tramite l'automazione.
    a. IAMAccess Analyzer può essere eseguito continuamente per generare avvisi quando le autorizzazioni possono essere potenzialmente ridotte.
4. Monitora e avvisa in caso di comportamenti anomali di accesso ai dati.
    a. GuardDutycontrolla sia le firme note delle minacce sia le deviazioni dai comportamenti di accesso di base per le risorse di archiviazione dei dati come EBS volumi, bucket S3 e database. RDS
5. Monitora e invia avvisi sui dati sensibili archiviati in luoghi inaspettati.
    a. Usa Amazon Macie per una scansione continua dei tuoi bucket S3 alla ricerca di dati sensibili.
6. Automatizza i backup sicuri e crittografati dei tuoi dati.
    a. AWS Backupè un servizio gestito che crea backup crittografati e sicuri di varie fonti di dati su. AWS Elastic Disaster Recovery consente di copiare carichi di lavoro completi sul server e mantenere una protezione continua dei dati con un obiettivo del punto di ripristino (RPO) misurato in secondi. È possibile configurare entrambi i servizi in modo che lavorino all'unisono per automatizzare la creazione di backup dei dati e la loro copia in posizioni di failover. Questo può aiutare a mantenere i dati disponibili in caso di eventi operativi o di sicurezza.

Risorse

Best practice correlate:

- SEC01-BP06 Automatizza l'implementazione dei controlli di sicurezza standard
- SEC03-BP02 Concedi l'accesso con privilegi minimi
- SEC03-BP04 Riduci continuamente le autorizzazioni
- SEC04-BP04 Avviare la riparazione delle risorse non conformi
- SEC07-BP03 Identificazione e classificazione automatizzate
- REL09-BP02 Backup sicuri e crittografati
- REL09-BP03 Esegue automaticamente il backup dei dati

Documenti correlati:

- AWS Guida prescrittiva: crittografa automaticamente i volumi Amazon esistenti e nuovi EBS
- Gestione del rischio di ransomware derivante dall' AWS utilizzo del NIST Cyber Security Framework () CSF

Esempi correlati:

- Come utilizzare regole e AWS CloudFormation hook AWS Config proattivi per impedire la creazione di risorse cloud non conformi
- Automatizza e gestisci centralmente la protezione dei dati per Amazon S3 con AWS Backup
- AWS re:Invent 2023 - Implementa la protezione proattiva dei dati utilizzando le istantanee di Amazon EBS
- AWS re:Invent 2022 - Build and automate for resilience with modern data protection

Strumenti correlati:

- AWS CloudFormation Guard
- AWS CloudFormation Guard Registro delle regole
- IAM Access Analyzer
- Amazon Macie
- AWS Backup
- Elastic Disaster Recovery

SEC08-BP04 Applica il controllo degli accessi

Per proteggere i dati a riposo, applica il controllo degli accessi utilizzando meccanismi come l'isolamento e il controllo delle versioni, quindi applica il principio del privilegio minimo. Impedisci l'accesso pubblico ai dati.

Risultato desiderato: verifica che solo gli utenti autorizzati possano accedere ai dati su base individuale. need-to-know La protezione dei dati è assicurata da backup regolari e dal controllo delle versioni, per evitare che la modifica dei dati o la loro eliminazione intenzionale o non voluta. L'isolamento dei dati critici dagli altri dati ne protegge la riservatezza e l'integrità.

Anti-pattern comuni:

- Archiviazione dei dati con requisiti di sensibilità o classificazione diversi.
- Utilizzo di autorizzazioni troppo permissive sulle chiavi di decrittografia.
- Classificazione impropria dei dati.
- Nessun mantenimento di backup dettagliati dei dati importanti.

- Accesso persistente ai dati di produzione.
- Nessun audit dell'accesso ai dati o revisione periodica delle autorizzazioni.

Livello di rischio associato se questa best practice non fosse adottata: basso

Guida all'implementazione

È possibile garantire la protezione dei dati a riposo mediante diversi controlli, tra cui l'accesso (utilizzando il privilegio minimo), l'isolamento e il controllo delle versioni. L'accesso ai dati deve essere verificato utilizzando meccanismi di rilevamento, come i log dei livelli di servizio AWS CloudTrail, come i log di accesso di Amazon Simple Storage Service (Amazon S3) Simple Storage Service (Amazon S3). Dovresti fare un inventario dei dati accessibili al pubblico e creare un piano per ridurre la quantità di dati disponibili al pubblico nel tempo.

Amazon S3 Glacier Vault Lock e Amazon S3 Object Lock forniscono un controllo di accesso obbligatorio per gli oggetti in Amazon S3: una volta bloccata con l'opzione di conformità, una policy di vault non può essere modificata nemmeno dall'utente root fino alla scadenza del blocco.

Passaggi dell'implementazione

- Applica il controllo degli accessi: applica il controllo degli accessi con privilegio minimo, incluso l'accesso alle chiavi di crittografia.
- Separa i dati in base a diversi livelli di classificazione: usa diversi Account AWS per i livelli di classificazione dei dati e gestisci questi account mediante AWS Organizations.
- Review AWS Key Management Service (AWS KMS): esamina il livello di accesso concesso nelle policy. AWS KMS
- Rivedi bucket Amazon S3 e autorizzazioni degli oggetti: rivedi con regolarità il livello di accesso concesso nelle policy dei bucket S3. La best practice è evitare di utilizzare bucket leggibili o scrivibili pubblicamente. Prendi in considerazione AWS Configl'idea di utilizzarlo per rilevare i bucket disponibili pubblicamente e Amazon CloudFront per distribuire contenuti da Amazon S3. Verifica che i bucket che non consentono l'accesso pubblico siano configurati correttamente per impedirlo. Per impostazione predefinita, tutti i bucket S3 sono privati e possono accedervi soltanto gli utenti a cui è stato esplicitamente accordato l'accesso.
- Usa AWS IAMAccess Analyzer: IAM Access Analyzer analizza i bucket Amazon S3 e genera un risultato quando una policy S3 concede l'accesso a un'entità esterna.
- Utilizza il controllo delle versioni di Amazon S3 e il blocco degli oggetti quando opportuno.

- Usa l'Inventario Amazon S3: l'Inventario Amazon S3 è uno degli strumenti che puoi utilizzare per eseguire audit e segnalare lo stato di replica e crittografia dei tuoi oggetti S3.
- Rivedi Amazon EBS e AMI le autorizzazioni di condivisione: le autorizzazioni di condivisione possono consentire la condivisione di immagini e volumi con soggetti Account AWS esterni al tuo carico di lavoro.
- Rivedi periodicamente le condivisioni di AWS Resource Access Manager per determinare se le risorse devono continuare a essere condivise. Resource Access Manager ti consente di condividere risorse, come policy AWS Network Firewall, regole resolver Amazon Route 53 e sottoreti, all'interno del tuo Amazon. VPCs Sottoponi regolarmente a audit le risorse condivise e interrompi la condivisione delle risorse che non devono più essere condivise.

Risorse

Best practice correlate:

- SEC03-BP01 Definire i requisiti di accesso
- SEC03-BP02 Concedi l'accesso con privilegi minimi

Documenti correlati:

- AWS KMS Whitepaper sui dettagli crittografici
- Introduzione alla gestione delle autorizzazioni di accesso alle risorse di Amazon S3
- Panoramica sulla gestione dell'accesso alle risorse AWS KMS
- Regole di AWS Config
- Amazon S3+ Amazon CloudFront: una combinazione creata nel cloud
- Utilizzo del controllo delle versioni
- Blocco degli oggetti mediante Object Lock di Amazon S3
- Condivisione di uno EBS snapshot Amazon
- Condiviso AMIs
- Ospitare un'applicazione a pagina singola su Amazon S3

Video correlati:

- Protezione dello storage a blocchi su AWS

SEC9. In che modo proteggi i dati in transito?

Proteggi i dati in transito implementando più controlli per ridurre il rischio di accessi non autorizzati o perdita.

Best practice

- SEC09-BP01 Implementare la gestione sicura di chiavi e certificati
- SEC09-BP02 Applica la crittografia in transito
- SEC09-BP03 Autentica le comunicazioni di rete

SEC09-BP01 Implementare la gestione sicura di chiavi e certificati

I certificati Transport Layer Security (TLS) vengono utilizzati per proteggere le comunicazioni di rete e stabilire l'identità di siti Web, risorse e carichi di lavoro su Internet, nonché su reti private.

Risultato desiderato: un sistema di gestione dei certificati sicuro in grado di fornire, distribuire, archiviare e rinnovare i certificati in un'infrastruttura a chiave pubblica (). PKI Un meccanismo sicuro di gestione delle chiavi e dei certificati impedisce la divulgazione del materiale relativo alle chiavi private dei certificati e rinnova in automatico il certificato su base periodica. Si integra inoltre con altri servizi per fornire comunicazioni di rete e identità sicure per le risorse delle macchine all'interno del carico di lavoro. Il materiale relativo alla chiave non dovrebbe mai essere accessibile alle identità umane.

Anti-pattern comuni:

- Esecuzione di passaggi manuali durante i processi di distribuzione, implementazione o rinnovo dei certificati.
- Attenzione insufficiente alla gerarchia delle autorità di certificazione (CA) durante la progettazione di una CA privata.
- Utilizzo di certificati autofirmati per risorse pubbliche.

Vantaggi dell'adozione di questa best practice:

- Semplificazione della gestione dei certificati attraverso la distribuzione, l'implementazione e il rinnovo automatizzati
- Incoraggia la crittografia dei dati in transito utilizzando i certificati TLS

- Maggiore sicurezza e verificabilità delle operazioni di certificazione intraprese dall'autorità di certificazione
- Organizzazione delle mansioni di gestione ai diversi livelli della gerarchia della CA

Livello di rischio associato se questa best practice non fosse adottata: elevato

Guida all'implementazione

I carichi di lavoro moderni fanno ampio uso di comunicazioni di rete crittografate utilizzando PKI protocolli comeTLS. PKILa gestione dei certificati può essere complessa, ma la fornitura, l'implementazione e il rinnovo automatizzati dei certificati possono ridurre le difficoltà associate alla gestione dei certificati.

AWS fornisce due servizi per gestire i PKI certificati generici: and (). AWS Certificate ManagerAWS Private Certificate AuthorityAWS Private CA ACMè il servizio principale utilizzato dai clienti per fornire, gestire e distribuire certificati da utilizzare sia in carichi di lavoro pubblici che privati. AWS ACMemette certificati utilizzando AWS Private CA e si integra con molti altri servizi AWS gestiti per fornire certificati sicuri per i carichi di lavoro. TLS

AWS Private CA consente di stabilire la propria autorità di certificazione principale o subordinata e di emettere TLS certificati tramite un. API È possibile utilizzare questo tipo di certificati in scenari in cui è possibile controllare e gestire la catena di fiducia sul lato client della TLS connessione. Oltre ai casi TLS d'uso, AWS Private CA possono essere utilizzati per emettere certificati su pod Kubernetes, attestazioni di prodotto dei dispositivi Matter, firmare codici e altri casi d'uso con un modello personalizzato. Puoi anche utilizzare IAMRoles Anywhere per fornire IAM credenziali temporanee ai carichi di lavoro locali a cui sono stati emessi certificati X.509 firmati dalla tua CA privata.

Oltre a ACM e AWS Private CA, AWS IoT Corefornisce supporto specializzato per il provisioning, la gestione e l'implementazione di PKI certificati su dispositivi IoT. AWS IoT Core fornisce meccanismi specializzati per l'onboarding dei dispositivi IoT nella tua infrastruttura a chiave pubblica su larga scala.

Considerazioni sulla creazione di una gerarchia CA privata

Quando occorre stabilire una CA privata, è importante prestare particolare attenzione a progettare in modo corretto la gerarchia della CA fin dall'inizio. È consigliabile distribuire ogni livello della gerarchia CA in modo separato Account AWS quando si crea una gerarchia CA privata. Questo passaggio intenzionale riduce la superficie di ogni livello della gerarchia delle CA, semplificando l'individuazione

delle anomalie nei dati di CloudTrail registro e riducendo l'ambito di accesso o l'impatto in caso di accesso non autorizzato a uno degli account. La CA principale deve risiedere in un account separato e va utilizzata solo per l'emissione di uno o più certificati CA intermedi.

Quindi, crea uno o più account CAs intermedi separati dall'account della CA principale per emettere certificati per utenti finali, dispositivi o altri carichi di lavoro. Infine, emetti certificati dalla tua CA principale a quella intermediaCAs, che a sua volta emetterà certificati agli utenti finali o ai dispositivi. Per ulteriori informazioni sulla pianificazione dell'implementazione della CA e sulla progettazione della gerarchia delle CA, inclusa la pianificazione della resilienza, la replica tra le regioni, la condivisione all'CAsinterno dell'organizzazione e altro ancora, consulta Pianificazione della distribuzione. AWS Private CA

Passaggi dell'implementazione

1. Determina i AWS servizi pertinenti richiesti per il tuo caso d'uso:
    - Molti casi d'uso possono sfruttare l'infrastruttura a chiave AWS pubblica esistente utilizzando AWS Certificate Manager. ACMpuò essere utilizzato per distribuire TLS certificati per server Web, sistemi di bilanciamento del carico o altri usi per certificati pubblicamente affidabili.
    - Prendi in considerazione AWS Private CA se occorre stabilire una gerarchia di autorità di certificazione privata o accedere a certificati esportabili. ACMpuò quindi essere utilizzato per emettere molti tipi di certificati di entità finale utilizzando. AWS Private CA
    - Per i casi d'uso in cui i certificati devono essere forniti su larga scala per dispositivi Internet delle cose (IoT) integrati, prendi in considerazione l'uso di AWS IoT Core.
2. Implementa il rinnovo automatico dei certificati quando possibile:
    - Utilizza il rinnovo ACM gestito per i certificati emessi da ACM insieme ai servizi AWS gestiti integrati.
3. Stabilisci la creazione di log e audit trail:
    - Abilita CloudTraili registri per tenere traccia degli accessi agli account che detengono le autorità di certificazione. Valuta la possibilità di configurare la convalida dell'integrità dei file di registro CloudTrail per verificare l'autenticità dei dati di registro.
    - Crea e rivedi a cadenza periodica report di audit che elencano i certificati emessi o revocati dalla tua CA privata. Questi report possono essere esportati in un bucket S3.
    - Quando si implementa una CA privata, è inoltre necessario creare un bucket S3 per archiviare l'elenco di revoca dei certificati (). CRL Per indicazioni sulla configurazione di questo bucket S3 in base ai requisiti del carico di lavoro, consulta Pianificazione di un elenco di revoca dei certificati (). CRL

Risorse

Best practice correlate:

- SEC02-BP02 Usa credenziali temporanee
- SEC08-BP01 Implementare una gestione sicura delle chiavi
- SEC09-BP03 Autentica le comunicazioni di rete

Documenti correlati:

- Come ospitare e gestire un'intera infrastruttura di certificati privata in AWS
- Come proteggere una gerarchia CA ACM privata su scala aziendale per il settore automobilistico e manifatturiero
- Private CA best practices
- Come utilizzarlo per AWS RAM condividere il tuo cross-account ACM Private CA

Video correlati:

- Activating AWS Certificate Manager Private CA (workshop)

Esempi correlati:

- Private CA workshop
- IOTWorkshop sulla gestione dei dispositivi (inclusa la fornitura dei dispositivi)

Strumenti correlati:

- Plugin per Kubernetes cert-manager da usare AWS Private CA

SEC09-BP02 Applica la crittografia in transito

Applica i requisiti di crittografia definiti in base alle policy, agli obblighi normativi e agli standard dell'organizzazione per contribuire a soddisfare i requisiti organizzativi, legali e di conformità. Utilizzate protocolli con crittografia solo quando trasmettete dati sensibili al di fuori del vostro cloud privato virtuale ()VPC. La crittografia aiuta a mantenere la riservatezza dei dati anche quando questi transitano su reti non affidabili.

Risultato desiderato: tutti i dati devono essere crittografati in transito utilizzando TLS protocolli sicuri e suite di crittografia. Il traffico di rete tra le tue risorse e Internet deve essere crittografato per evitare l'accesso non autorizzato ai dati. Il traffico di rete esclusivamente all'interno AWS dell'ambiente interno deve essere crittografato TLS ove possibile. La rete AWS interna è crittografata per impostazione predefinita e il traffico di rete all'interno di un VPC non può essere falsificato o sniffato a meno che una parte non autorizzata non abbia ottenuto l'accesso a qualsiasi risorsa che genera traffico (come istanze EC2 Amazon e contenitori Amazon). ECS Prendi in considerazione la possibilità di proteggere il network-to-network traffico con una rete privata virtuale (). IPsec VPN

Anti-pattern comuni:

- Utilizzo di versioni obsolete e componenti della suite di crittografia (ad esempio SSLTLS, SSL v3.0, chiavi a RSA 1024 bit e cipher). RC4
- Consentire il traffico non crittografato () da o verso risorse rivolte al pubblico. HTTP
- Monitoraggio e sostituzione mancati dei certificati X.509 prima della scadenza.
- Utilizzo di certificati X.509 autofirmati per. TLS

Livello di rischio associato se questa best practice non fosse adottata: elevato

Guida all'implementazione

AWS i servizi forniscono HTTPS endpoint utilizzati TLS per la comunicazione, fornendo la crittografia in transito durante la comunicazione con. AWS APIs I protocolli non sicuri, ad esempio, HTTP possono essere controllati e bloccati in un attimo VPC tramite l'uso di gruppi di sicurezza. HTTPIe richieste possono anche essere reindirizzate automaticamente su Amazon CloudFront o su un Application Load Balancer. HTTPS Hai il controllo completo sulle tue risorse informatiche per implementare la crittografia in transito nei tuoi servizi. Inoltre, puoi utilizzare la VPN connettività VPC proveniente da una rete esterna o AWS Direct Connectper facilitare la crittografia del traffico. Verifica che i tuoi clienti stiano effettuando chiamate AWS APIs utilizzando almeno la versione TLS 1.2, così come l'uso delle versioni precedenti di giugno 2023 AWS è obsoleto TLS. AWS consiglia di utilizzare 1.3. TLS Le soluzioni di terze parti sono disponibili in Marketplace AWS caso di esigenze particolari.

Passaggi dell'implementazione

- Applica la crittografia in transito: i requisiti di crittografia definiti dovrebbero essere basati sugli standard e sulle best practice più recenti e consentire solo protocolli sicuri. Ad esempio, configura un gruppo di sicurezza per consentire il HTTPS protocollo solo a un sistema di bilanciamento del carico delle applicazioni o a un'EC2istanza Amazon.

- Configura protocolli sicuri nei servizi edge: configura HTTPS con Amazon CloudFront e usa un profilo di sicurezza appropriato al tuo livello di sicurezza e al tuo caso d'uso.
- Usa un VPNper la connettività esterna: prendi in considerazione l'utilizzo di un IPsec VPN per proteggere le point-to-point network-to-network connessioni per garantire la privacy e l'integrità dei dati.
- Configura protocolli sicuri nei bilanciatori del carico: seleziona una policy di sicurezza che fornisca le suite di crittografia più solide supportate dai client che si connetteranno all'ascoltatore. Crea un HTTPS listener per il tuo Application Load Balancer.
- Configura protocolli sicuri in Amazon Redshift: configura il cluster per richiedere una connessione Secure Socket Layer (SSL) o Transport Layer Security (TLS).
- Configura protocolli sicuri: consulta la documentazione AWS del servizio per determinare encryption-in-transit le funzionalità.
- Configura l'accesso sicuro durante il caricamento su bucket Amazon S3: utilizza i controlli delle policy sui bucket Amazon S3 per applicare l'accesso sicuro ai dati.
- Prendi in considerazione l'utilizzo AWS Certificate Manager: ACM consente di fornire, gestire e distribuire TLS certificati pubblici da utilizzare con AWS i servizi.
- Prendi in considerazione l'utilizzo AWS Private Certificate Authorityper PKI esigenze private: AWS Private CA consente di creare gerarchie di autorità di certificazione (CA) private per emettere certificati X.509 di entità finale che possono essere utilizzati per creare canali crittografati. TLS

Risorse

Documenti correlati:

- HTTPSUtilizzo con CloudFront
- Connect il tuo VPC a reti remote utilizzando AWS Virtual Private Network
- Crea un HTTPS listener per il tuo Application Load Balancer
- Tutorial: ConfigurazioneSSL/TLSsu Amazon Linux 2
- Utilizzo TLS diSSL/per crittografare una connessione a un'istanza DB
- Configuring security options for connections

SEC09-BP03 Autentica le comunicazioni di rete

Verifica l'identità delle comunicazioni utilizzando protocolli che supportano l'autenticazione, come Transport Layer Security (TLS) oIPsec.

Progetta il carico di lavoro in modo da utilizzare protocolli di rete sicuri e autenticati per le comunicazioni tra servizi, applicazioni o utenti. L'utilizzo di protocolli di rete che supportano autenticazione e autorizzazione offre un controllo più rigido sui flussi di rete e riduce l'impatto di eventuali accessi non autorizzati.

Risultato desiderato: un carico di lavoro con un piano dati ben definito e flussi di traffico del piano di controllo (control-plane) tra i servizi. I flussi di traffico utilizzano protocolli di rete autenticati e crittografati laddove tecnicamente fattibile.

Anti-pattern comuni:

- Flussi di traffico non crittografati o non autenticati all'interno del carico di lavoro.
- Riutilizzo delle credenziali di autenticazione tra più utenti o entità.
- Uso esclusivo di controlli di rete come meccanismo di controllo degli accessi.
- Creazione di un meccanismo di autenticazione personalizzato anziché usare meccanismi di autenticazione standard del settore.
- Flussi di traffico eccessivamente permissivi tra i componenti del servizio o altre risorse di. VPC

Vantaggi dell'adozione di questa best practice:

- Limita l'ambito dell'impatto di eventuali accessi non autorizzati a una parte del carico di lavoro.
- Fornisce un livello maggiore di sicurezza affinché solo entità autenticate eseguano le azioni.
- Migliora il disaccoppiamento dei servizi definendo e applicando in modo chiaro le interfacce di trasferimento dei dati previste.
- Migliora monitoraggio, creazione di log e risposta agli incidenti tramite l'attribuzione di richieste e interfacce di comunicazione ben definite.
- Gestisce defense-in-depth i carichi di lavoro combinando i controlli di rete con i controlli di autenticazione e autorizzazione.

Livello di rischio associato se questa best practice non fosse adottata: basso

Guida all'implementazione

È possibile suddividere i modelli di traffico di rete del tuo carico di lavoro in due categorie:

- Il traffico est-ovest corrisponde ai flussi di traffico tra i servizi facenti parte di un carico di lavoro.

- Il traffico nord-sud rappresenta i flussi di traffico tra carico di lavoro e consumatori.

Sebbene crittografare il traffico nord-sud sia la prassi comune, proteggere il traffico est-ovest mediante protocolli autenticati non è così frequente. Le moderne best practice di sicurezza raccomandano che la progettazione della rete non sia l'unico elemento in grado di garantire una relazione affidabile tra due entità. Quando due servizi possono trovarsi all'interno di una rete comune, è comunque consigliabile crittografare, autenticare e autorizzare le comunicazioni tra tali servizi.

Ad esempio, il AWS servizio APIs utilizza il protocollo di AWS firma Signature Version 4 (SigV4) per autenticare il chiamante, indipendentemente dalla rete da cui proviene la richiesta. Questa autenticazione garantisce che sia AWS APIs possibile verificare l'identità che ha richiesto l'azione e che tale identità possa quindi essere combinata con le politiche per prendere una decisione di autorizzazione per determinare se l'azione debba essere consentita o meno.

Servizi come Amazon VPC Lattice e Amazon API Gateway consentono di utilizzare lo stesso protocollo di firma SigV4 per aggiungere autenticazione e autorizzazione al traffico est-ovest nei propri carichi di lavoro. Se le risorse esterne all' AWS ambiente devono comunicare con servizi che richiedono l'autenticazione e l'autorizzazione basate su SigV4, puoi utilizzare AWS Identity and Access Management (IAM) Roles Anywhere sulla risorsa non utilizzata per acquisire credenziali temporanee.AWS AWS Queste credenziali possono essere utilizzate per firmare richieste ai servizi che utilizzano SigV4 per autorizzare l'accesso.

Un altro meccanismo comune per l'autenticazione del traffico est-ovest è l'autenticazione reciproca (m). TLS TLS Molte business-to-business applicazioni, microservizi e Internet of Things (IoT) utilizzano m TLS per convalidare l'identità di entrambi i lati di una TLS comunicazione mediante l'uso di certificati X.509 lato client e lato server. Questi certificati possono essere emessi da (). AWS Private Certificate Authority AWS Private CA Puoi utilizzare servizi come Amazon API Gateway e AWS App Meshfornire TLS l'autenticazione m per le comunicazioni tra o tra carichi di lavoro. Sebbene m TLS fornisca informazioni di autenticazione per entrambi i lati di una TLS comunicazione, non fornisce un meccanismo di autorizzazione.

Infine, OAuth 2.0 e OpenID Connect (OIDC) sono due protocolli tipicamente utilizzati per controllare l'accesso ai servizi da parte degli utenti, ma ora stanno diventando popolari anche per il service-to-service traffico. APIGateway fornisce un autorizzatore JSON Web Token (JWT), che consente ai carichi di lavoro di limitare l'accesso ai API percorsi utilizzando provider di identità JWTs emessi da OIDC o OAuth 2.0. OAuth2gli ambiti possono essere utilizzati come fonte per le decisioni di autorizzazione di base, ma i controlli di autorizzazione devono ancora essere implementati a livello

di applicazione e gli OAuth2 ambiti da soli non possono supportare esigenze di autorizzazione più complesse.

Passaggi dell'implementazione

- Definisci e documenta i flussi di rete del carico di lavoro: il primo passo nell'implementazione di una defense-in-depth strategia è definire i flussi di traffico del carico di lavoro.
  - Crea un diagramma del flusso di dati che definisca in modo chiaro le modalità di trasmissione dei dati tra i diversi servizi che costituiscono il carico di lavoro. Questo diagramma è il primo passo per autorizzare tali flussi nei canali di rete autenticati.
  - Nelle fasi di sviluppo e test dota il carico di lavoro di strumenti per controllare che il diagramma del flusso di dati rifletta in modo preciso il comportamento del carico di lavoro in fase di runtime.
  - Un diagramma del flusso di dati può essere utile anche quando si esegue un esercizio di modellazione delle minacce, come descritto in SEC01-BP07 Identificare le minacce e assegnare priorità alle mitigazioni utilizzando un modello di minaccia.
- Stabilisci i controlli di rete: valuta la possibilità di stabilire controlli di rete allineati ai AWS flussi di dati. Sebbene i confini di rete non debbano essere l'unico controllo di sicurezza, forniscono un livello di defense-in-depth strategia per proteggere il carico di lavoro.
  - Utilizza i gruppi di sicurezza per stabilire, definire e limitare i flussi di dati tra le risorse.
  - Valuta la possibilità AWS PrivateLinkdi utilizzarlo per comunicare sia con AWS i servizi di terze parti che lo supportano AWS PrivateLink. I dati inviati tramite un endpoint di AWS PrivateLink interfaccia rimangono all'interno della spina dorsale della AWS rete e non attraversano la rete Internet pubblica.
- Implementa l'autenticazione e l'autorizzazione tra i servizi del tuo carico di lavoro: scegli il set di AWS servizi più appropriato per fornire flussi di traffico autenticati e crittografati nel tuo carico di lavoro.
  - Prendi in considerazione Amazon VPC Lattice per una service-to-service comunicazione sicura. VPCLattice può utilizzare l'autenticazione SigV4 combinata con le politiche di autenticazione per controllare l'accesso. service-to-service
  - Per la service-to-service comunicazione tramite mTLS, considera APIGateway o App Mesh. AWS Private CApuò essere utilizzato per stabilire una gerarchia CA privata in grado di emettere certificati da utilizzare con m. TLS
  - Durante l'integrazione con servizi che utilizzano OAuth 2.0 oppureOIDC, considera APIGateway che utilizza l'autorizzatore. JWT

    - Per la comunicazione tra il carico di lavoro e i dispositivi IoT, prendi in considerazione AWS IoT Core, che offre diverse opzioni per la crittografia e l'autenticazione del traffico di rete.
- Monitora gli accessi non autorizzati: monitora in modo continuo i canali di comunicazione non intenzionali, i tentativi di accesso dei principali non autorizzati a risorse protette e altri schemi di accesso impropri.
    - Se utilizzi VPC Lattice per gestire l'accesso ai tuoi servizi, prendi in considerazione l'abilitazione e il monitoraggio dei log di accesso di VPCLattice. Questi registri di accesso includono informazioni sull'entità richiedente, informazioni di rete tra cui origine e destinazione VPC e metadati della richiesta.
    - Valuta la possibilità di abilitare i log di VPC flusso per acquisire i metadati sui flussi di rete e verificare periodicamente la presenza di anomalie.
    - Consulta la AWS Security Incident Response Guide e la sezione Incident Response del pilastro di sicurezza AWS Well-Architected Framework per ulteriori indicazioni sulla pianificazione, la simulazione e la risposta agli incidenti di sicurezza.

Risorse

Best practice correlate:

- SEC03-BP07 Analizza l'accesso pubblico e tra account
- SEC02-BP02 Usa credenziali temporanee
- SEC01-BP07 Identifica le minacce e dai priorità alle mitigazioni utilizzando un modello di minaccia

Documenti correlati:

- Valutazione dei metodi di controllo degli accessi per proteggere Amazon API Gateway APIs
- Configurazione dell'TLSautenticazione reciproca per un REST API
- Come proteggere gli HTTP endpoint API Gateway con un sistema di autorizzazione JWT
- Autorizzazione delle chiamate dirette ai AWS servizi utilizzando il fornitore di credenziali AWS IoT Core
- AWS Guida alla risposta agli incidenti di sicurezza

Video correlati:

- AWS re:invent 2022: Presentazione di Lattice VPC

- AWS re:invent 2020: autenticazione serverless per on API HTTP APIs AWS

Esempi correlati:

- Workshop Amazon VPC Lattice
- Zero-Trust Episode 1 – The Phantom Service Perimeter workshop

Risposta agli incidenti

Domanda

- SEC10. In che modo è possibile prevedere gli incidenti, rispondere agli stessi e risolverli?

SEC10. In che modo è possibile prevedere gli incidenti, rispondere agli stessi e risolverli?

Anche se dispone di controlli preventivi e di rilevamento maturi, l'organizzazione deve ancora implementare meccanismi per rispondere e mitigare il potenziale impatto degli incidenti di sicurezza. La tua preparazione influisce fortemente sulla capacità dei team di operare in modo efficace durante un incidente, isolare, contenere ed eseguire indagini sui problemi e ripristinare le operazioni a uno stato valido noto. La messa in atto degli strumenti e l'accesso prima di un incidente di sicurezza, quindi la pratica sistematica della risposta agli incidenti durante le giornate di gioco, aiuterà a garantire il ripristino, riducendo al minimo le interruzioni dell'attività.

Best practice

- SEC10-BP01 Identifica il personale chiave e le risorse esterne
- SEC10-BP02 Sviluppare piani di gestione degli incidenti
- SEC10-BP03 Preparare le capacità forensi
- SEC10-BP04 Sviluppare e testare i playbook di risposta agli incidenti di sicurezza
- SEC10-BP05 Accesso preliminare alla fornitura
- SEC10-BP06 Strumenti di pre-installazione
- SEC10-BP07 Esegui simulazioni
- SEC10-BP08 Stabilire un framework per imparare dagli incidenti

SEC10-BP01 Identifica il personale chiave e le risorse esterne

Identifica personale, risorse e requisiti legali interni ed esterni per consentire all'organizzazione a rispondere a un incidente.

Risultato desiderato: presenza di un elenco del personale chiave, delle relative informazioni di contatto e dei ruoli svolti nel rispondere a un evento di sicurezza. Rivedi queste informazioni con regolarità e aggiornarle per riflettere i cambiamenti del personale dal punto di vista degli strumenti interni ed esterni. Nel documentare queste informazioni, prendete in considerazione tutti i fornitori e i fornitori di servizi di terze parti, inclusi i partner di sicurezza, i fornitori di servizi cloud e le applicazioni (software-as-a-serviceSaaS). Durante un evento di sicurezza, il personale è disponibile con il livello di responsabilità, il contesto e l'accesso appropriati per poter rispondere ed eseguire il ripristino.

Anti-pattern comuni:

- Mancata tenuta di un elenco aggiornato del personale chiave con le informazioni di contatto, i ruoli e le responsabilità in caso di risposta a eventi di sicurezza.
- Si presume che tutti conoscano persone, dipendenze, infrastruttura e soluzioni per rispondere a un evento ed eseguire il ripristino dopo lo stesso.
- Mancata predisposizione di un archivio di documenti o conoscenze che rappresenti l'infrastruttura o la progettazione di applicazioni chiave.
- Mancata predisposizione di processi di onboarding adeguati per i nuovi dipendenti, in modo che possano contribuire in modo efficace alla risposta a un evento di sicurezza, come la realizzazione di simulazioni di eventi.
- Mancata predisposizione di un percorso di escalation quando il personale chiave è temporaneamente non disponibile o non risponde durante gli eventi di sicurezza.

Vantaggi dell'adozione di questa best practice: riduzione del tempo di valutazione e risposta impiegato per identificare il personale giusto e il relativo ruolo durante un evento grazie a questa pratica. Riduci al minimo le perdite di tempo durante un evento mantenendo un elenco aggiornato del personale chiave e dei relativi ruoli, in modo da poter portare le persone giuste al triage e al ripristino da un evento.

Livello di rischio associato se questa best practice non fosse adottata: elevato

Guida all'implementazione

Identifica il personale chiave all'interno dell'organizzazione: conserva un elenco di contatti del personale interno alla tua organizzazione che potrebbe essere necessario coinvolgere. Rivedi e aggiorna in modo regolare queste informazioni in caso di spostamento del personale, quali modifiche organizzative, promozioni e cambi di team. Questo è particolarmente importante per i ruoli chiave come gli incident manager, i team di risposta e i responsabili delle comunicazioni.

- Responsabile degli incidenti: i responsabili degli incidenti dispongono dell'autorità generale durante la risposta all'evento.
- Persone che intervengono dopo un incidente: le persone che intervengono dopo un incidente sono responsabili delle attività di indagine e correzione. Queste persone possono differire in base al tipo di evento, ma in genere sono sviluppatori e team operativi responsabili dell'applicazione interessata.
- Responsabile delle comunicazioni: il responsabile delle comunicazioni gestisce comunicazioni interne ed esterne, in particolare con gli enti pubblici, le autorità di regolamentazione e i clienti.
- Esperti in materia (SMEs): nel caso di team distribuiti e autonomi, ti consigliamo di identificarne uno SME per carichi di lavoro mission critical. Queste persone offrono approfondimenti su funzionamento e classificazione dei dati dei carichi di lavoro critici coinvolti nell'evento.

Prendi in considerazione l'utilizzo della funzionalità AWS Systems Manager Incident Manager per l'acquisizione dei contatti chiave, la definizione di un piano di risposta, l'automazione degli orari delle chiamate e la creazione di piani di escalation. Automatizza e organizza i turni per tutto il personale attraverso un programma di chiamata, in modo che la responsabilità del carico di lavoro sia condivisa tra i proprietari. Ciò promuove buone pratiche, come l'emissione di metriche e log pertinenti e la definizione di soglie di allarme importanti per il carico di lavoro.

Identifica i partner esterni: le aziende utilizzano strumenti creati da fornitori di software indipendenti (ISVs), partner e subappaltatori per creare soluzioni differenziate per i propri clienti. Coinvolgi il personale chiave di queste parti che può aiutarti a rispondere e a eseguire il ripristino dopo un incidente. Ti consigliamo di iscriverti al livello appropriato per ottenere un rapido accesso agli esperti AWS Support in AWS materia tramite un caso di supporto. Prendi in considerazione accordi simili con tutti i fornitori di soluzioni critiche per i carichi di lavoro. Alcuni eventi di sicurezza richiedono alle aziende quotate in borsa di notificare evento ed effetti agli enti pubblici e alle autorità di regolamentazione pertinenti. Mantieni e aggiorna le informazioni di contatto per i dipartimenti pertinenti e le persone responsabili.

Passaggi dell'implementazione

1. Configura una soluzione per la gestione degli incidenti.
   a. Prendi in considerazione l'implementazione di Incident Manager nel tuo account Security Tooling.
2. Definisci i contatti nella tua soluzione di gestione degli incidenti.
   a. Definisci almeno due tipi di canali di contatto per ogni contatto (ad SMS esempio telefono o e-mail), per garantire la raggiungibilità durante un incidente.
3. Definisci un piano di risposta.
   a. Identifica i contatti più opportuni da coinvolgere durante un incidente. Definisci piani di escalation in linea con i ruoli del personale da coinvolgere, piuttosto che con i singoli contatti. Valuta la possibilità di includere i contatti che potrebbero essere responsabili dell'informare entità esterne, anche se non direttamente coinvolti nella risoluzione dell'incidente.

Risorse

Best practice correlate:

- OPS02-BP03 Le attività operative hanno identificato i proprietari responsabili delle loro prestazioni

Documenti correlati:

- AWS Security Incident Response Guide

Esempi correlati:

- Framework AWS per playbook per i clienti
- Prepare for and respond to security incidents in your AWS environment

Strumenti correlati:

- AWS Systems Manager Incident Manager

Video correlati:

- Amazon's approach to security during development

SEC10-BP02 Sviluppare piani di gestione degli incidenti

Il primo documento da predisporre per la risposta agli incidenti è il piano di risposta agli incidenti. Lo scopo del piano di risposta agli incidenti è costituire la base del programma e della strategia di risposta agli incidenti.

Vantaggi dell'adozione di questa best practice: lo sviluppo di processi di risposta agli incidenti completi e definiti in modo chiaro è fondamentale per un programma di risposta agli incidenti efficace e scalabile. Quando si verifica un evento di sicurezza, passaggi e flussi di lavoro ben definiti agevoleranno una risposta tempestiva. Potrebbero essere già presenti processi di risposta agli incidenti. Indipendentemente dallo stato attuale, è importante aggiornare, iterare e testare con regolarità i processi di risposta agli incidenti.

Livello di rischio associato se questa best practice non fosse adottata: elevato

Guida all'implementazione

Un piano di gestione degli incidenti è fondamentale per rispondere, mitigare ed eseguire il ripristino a seguito del potenziale impatto degli incidenti di sicurezza. Un piano di gestione degli incidenti è un processo strutturato volto a identificare, correggere e rispondere tempestivamente agli incidenti di sicurezza.

Il cloud presenta molti degli stessi ruoli e requisiti operativi che si trovano in un ambiente on-premises. Nella creazione di un piano di gestione degli incidenti, è importante tenere conto delle strategie di risposta e ripristino ideali per i risultati aziendali e ai requisiti di conformità. Ad esempio, se gestisci carichi di lavoro RAMP conformi alla Fed negli AWS Stati Uniti, è utile attenersi alla SP 800-61 Computer Security Handling Guide. NIST Allo stesso modo, quando gestisci carichi di lavoro con dati europei di identificazione personale (PII), prendi in considerazione scenari come proteggere e rispondere ai problemi relativi alla residenza dei dati, come previsto dal Regolamento generale sulla protezione dei dati () dell'UE. GDPR

Quando crei un piano di gestione degli incidenti per i tuoi carichi di lavoro AWS, inizia con il modello di responsabilitàAWS condivisa per creare un defense-in-depth approccio alla risposta agli incidenti. In questo modello, AWS gestisce la sicurezza del cloud e tu sei responsabile della sicurezza nel cloud. Ciò significa che mantieni il controllo e sei responsabile dei controlli di sicurezza che scegli di implementare. La AWS Security Incident Response Guide illustra concetti chiave e linee guida di base per la creazione di un piano di gestione degli incidenti incentrato sul cloud.

Un piano di gestione degli incidenti efficace va iterato in modo continuo per rimanere in linea con l'obiettivo delle operazioni cloud. Prendi in considerazione l'utilizzo dei piani di implementazione illustrati di seguito durante la creazione e l'evoluzione del tuo piano di gestione degli incidenti.

Passaggi dell'implementazione

Definisci ruoli e responsabilità

La gestione degli eventi di sicurezza richiede disciplina interorganizzativa e propensione all'azione. All'interno della struttura organizzativa, dovrebbero esserci molte persone da considerarsi responsabili, affidabili, consultabili o informate durante un incidente, come i rappresentanti delle risorse umane (HR), i membri del team esecutivo e quelli dell'ufficio legale. Considera questi ruoli e queste responsabilità e, se è necessario, coinvolgi terze parti. Si noti che molte aree geografiche presentano leggi locali che disciplinano ciò che è consentito e ciò che non lo è. Sebbene possa sembrare burocratico creare uno schema responsabile, affidabile, consultato e informato (RACI) per i piani di risposta alla sicurezza, così facendo si facilita una comunicazione rapida e diretta e si delinea chiaramente la leadership nelle diverse fasi dell'evento.

Durante un incidente, coinvolgere i proprietari e gli sviluppatori delle applicazioni e delle risorse interessate è fondamentale perché si tratta di esperti in materia (SMEs) in grado di fornire informazioni e contesto per contribuire alla misurazione dell'impatto. Assicurati di fare pratica e instaurare relazioni con sviluppatori e proprietari delle applicazioni prima di affidarti alla loro esperienza per la gestione della risposta agli incidenti. I proprietari delle applicazioni oSMEs, ad esempio, gli amministratori o gli ingegneri del cloud, potrebbero dover agire in situazioni in cui l'ambiente non è familiare o presenta complessità o in cui i soccorritori non hanno accesso.

Infine, nell'indagine o nella risposta potrebbero essere coinvolti partner affidabili vista la loro capacità di fornire competenze aggiuntive e capacità analitiche strategiche. Quando non disponi di queste competenze nel tuo team, potresti voler assumere una persona esterna per assistenza.

Comprendi i team di AWS risposta e il supporto

- AWS Support
  - AWS Supportoffre una gamma di piani che forniscono l'accesso a strumenti e competenze che supportano il successo e lo stato operativo delle vostre AWS soluzioni. Se avete bisogno di supporto tecnico e di ulteriori risorse per pianificare, implementare e ottimizzare AWS l'ambiente, potete selezionare il piano di supporto più adatto al vostro caso AWS d'uso.
  - Considera il Support Center in AWS Management Console (è richiesto l'accesso) come punto di contatto centrale per ricevere assistenza per problemi che riguardano AWS le tue risorse.

L'accesso a AWS Support è controllato da AWS Identity and Access Management. Per ulteriori informazioni su come accedere alle AWS Support funzionalità, consulta Guida introduttiva AWS Support.

- AWS Customer Incident Response Team (CIRT)
  - Il AWS Customer Incident Response Team (CIRT) è un AWS team globale specializzato 24 ore su 24, 7 giorni su 7, che fornisce supporto ai clienti durante gli eventi di sicurezza attivi dal punto di vista del cliente nell'ambito del Modello di responsabilitàAWS condivisa.
  - Quando ti AWS CIRT supporta, fornisce assistenza nella valutazione e nel ripristino di un evento di sicurezza attivo. AWS Possono fornire assistenza nell'analisi delle cause principali tramite l'uso dei log di AWS servizio e fornire consigli per il ripristino. Può altresì fornire consigli e best practice sulla sicurezza così da evitare eventi di sicurezza in futuro.
  - AWS i clienti possono coinvolgerli AWS CIRT attraverso un AWS Support caso.
- DDoSsupporto alla risposta
  - AWS offre AWS Shield, che fornisce un servizio di protezione gestito dalla denial of service (DDoS) distribuito che salvaguarda le applicazioni Web in esecuzione su. AWS Shield offre un rilevamento sempre attivo e mitigazioni automatiche in linea che possono ridurre al minimo i tempi di inattività e la latenza delle applicazioni, quindi non è necessario impegnarsi per trarre vantaggio dalla protezione. AWS Support DDoS Esistono due livelli di Shield: AWS Shield Standard e AWS Shield Advanced. Per maggiori informazioni sulle differenze tra questi due livelli, consulta la documentazione della funzionalità Shield.
- AWS Managed Services (AMS)
  - AWS Managed Services (AMS) fornisce una gestione continua dell' AWS infrastruttura in modo da potersi concentrare sulle applicazioni. Implementando le migliori pratiche per la manutenzione dell'infrastruttura, AMS contribuisce a ridurre i costi operativi e i rischi. AMSautomatizza attività comuni come richieste di modifica, monitoraggio, gestione delle patch, sicurezza e servizi di backup e fornisce servizi per l'intero ciclo di vita per la fornitura, l'esecuzione e il supporto dell'infrastruttura.
  - AMSsi assume la responsabilità dell'implementazione di una suite di controlli di sicurezza e fornisce una prima linea di risposta agli avvisi 24 ore su 24, 7 giorni su 7. Quando viene avviato un avviso, AMS segue una serie standard di playbook automatici e manuali per verificare una risposta coerente. Questi playbook vengono condivisi con AMS i clienti durante l'onboarding in modo che possano sviluppare e coordinare una risposta. AMS

Sviluppo di piani di risposta agli incidenti

Lo scopo del piano di risposta agli incidenti è costituire la base del programma e della strategia di risposta agli incidenti. Il piano di risposta agli incidenti deve essere contenuto in un documento formale. Un piano di risposta agli incidenti include in genere le seguenti sezioni:

- Una panoramica del team di risposta agli incidenti: delinea obiettivi e funzioni del team di risposta agli incidenti.
- Ruoli e responsabilità: indica le parti interessate alla risposta agli incidenti e illustra in dettaglio i loro ruoli in caso di incidente.
- Un piano di comunicazione: fornisce dettagli sulle informazioni di contatto e sulle tue modalità di comunicazione durante un incidente.
- Metodi di comunicazione di backup: è consigliabile disporre della out-of-band comunicazione come backup per la comunicazione degli incidenti. Un esempio di applicazione che fornisce un canale di out-of-band comunicazione sicuro è AWS Wickr.
- Fasi di risposta agli incidenti e azioni da intraprendere: elenca le fasi della risposta agli incidenti (ad esempio, rilevamento, analisi, eliminazione, contenimento e ripristino), comprese le azioni di alto livello da intraprendere all'interno di tali fasi.
- Definizioni di gravità e assegnazione della priorità agli incidenti: illustra in dettaglio come classificare la gravità di un incidente, le modalità di assegnazione della priorità all'incidente e, quindi, in che modo le definizioni di gravità influiscono sulle procedure di escalation.

Sebbene queste sezioni siano comuni ad aziende di diverse dimensioni e settori, il piano di risposta agli incidenti di ciascuna organizzazione è unico. Devi creare un piano di risposta agli incidenti che funzioni al meglio per la tua organizzazione.

Risorse

Best practice correlate:

- SEC04 (Come si rilevano e analizzano gli eventi di sicurezza?)

Documenti correlati:

- AWS Guida alla risposta agli incidenti di sicurezza
- NIST: Guida alla gestione degli incidenti di sicurezza informatica

SEC10-BP03 Preparare le capacità forensi

Prima che si verifichi un incidente di sicurezza, puoi sviluppare funzionalità forensi per supportare le indagini sugli eventi di sicurezza.

Livello di rischio associato se questa best practice non fosse adottata: medio

Si applicano i concetti della tradizionale analisi forense locale a. AWS Per informazioni chiave su cui iniziare a sviluppare funzionalità di analisi forense in Cloud AWS, vedere Strategie dell'ambiente di indagine forense in. Cloud AWS

Una volta configurati l'ambiente e la Account AWS struttura per le indagini forensi, definisci le tecnologie necessarie per eseguire efficacemente metodologie valide dal punto di vista forense nelle quattro fasi:

- Raccolta: raccogli i AWS log pertinenti, ad esempio i log di VPC flusso e i log a AWS CloudTrail livello di AWS Config host. Raccogli istantanee, backup e dump di memoria delle risorse interessate, ove disponibili. AWS
- Esame: rivedi i dati raccolti estraendo e valutando le informazioni pertinenti.
- Analisi: analizza i dati raccolti per comprendere l'incidente e trarre le conclusioni.
- Creazione di report: presenta le informazioni risultanti dalla fase di analisi.

Passaggi dell'implementazione

Preparazione dell'ambiente per le funzionalità forensi

AWS Organizationsti aiuta a gestire e governare centralmente un AWS ambiente man mano che cresci e scalerai le risorse. AWS Un' AWS organizzazione consolida le tue Account AWS in modo che tu possa amministrarle come una singola unità. È possibile utilizzare le unità organizzative (OUs) per raggruppare gli account e amministrarli come un'unica unità.

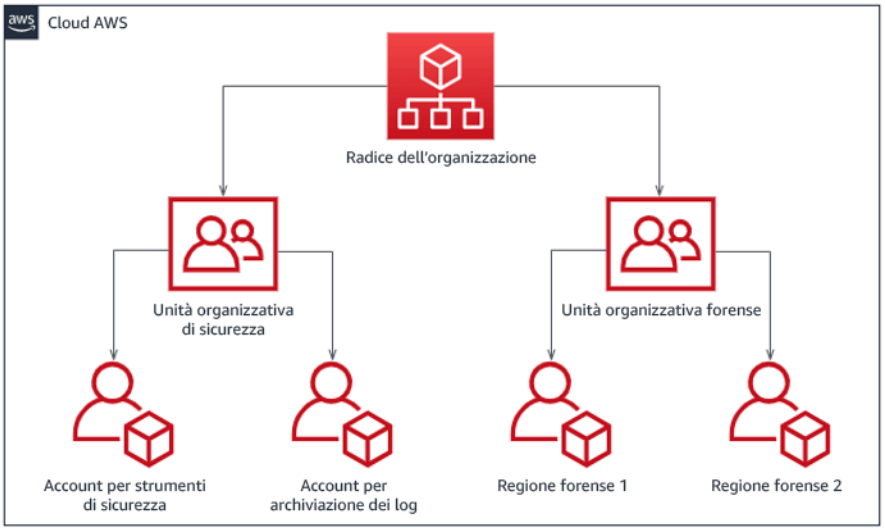
Per la risposta agli incidenti, è utile disporre di una Account AWS struttura che supporti le funzioni di risposta agli incidenti, che includa un'unità organizzativa di sicurezza e un'unità organizzativa forense. All'interno dell'unità organizzativa di sicurezza, è necessario disporre degli account per:

- Archiviazione dei log: aggrega i log in un archivio di log con autorizzazioni limitate. Account AWS
- Strumenti di sicurezza: centralizza i servizi di sicurezza in uno strumento di sicurezza. Account AWS Questo account funge da amministratore delegato per i servizi di sicurezza.

Nell'unità organizzativa con funzionalità forensi, puoi implementare uno o più account con funzionalità forensi per ciascuna regione in cui operi, a seconda di quale è più adatta all'azienda e al modello operativo. Se crei un account forense per regione, puoi bloccare la creazione di AWS risorse al di fuori di tale regione e ridurre il rischio che le risorse vengano copiate in un'area non prevista. Ad esempio, se operi solo nella regione degli Stati Uniti orientali (Virginia settentrionale) (`us-east-1`) e Stati Uniti occidentali (Oregon) (`us-west-2`), nell'unità organizzativa con funzionalità forensi avrai due account: uno per `us-east-1` e uno per `us-west-2`.

È possibile creare un'analisi Account AWS forense per più regioni. Dovresti prestare attenzione nel copiare AWS le risorse su quell'account per verificare che tu stia rispettando i requisiti di sovranità dei dati. Poiché la creazione di nuovi account richiede tempo, è fondamentale creare e fornire gli strumenti adatti agli account con funzionalità forensi con largo anticipo rispetto agli incidenti, in modo che gli addetti siano preparati a utilizzarli in modo efficace per la risposta.

Il diagramma seguente mostra una struttura degli account di esempio che include un'unità organizzativa con funzionalità forensi con account con funzionalità forensi per regione:



Struttura degli account per regione per la risposta agli incidenti

Acquisizione di backup e snapshot

La configurazione dei backup di sistemi e database importanti è fondamentale per il ripristino da un incidente di sicurezza e per scopi forensi. Grazie ai backup puoi ripristinare i tuoi sistemi allo stato di sicurezza precedente. Sì AWS, puoi scattare istantanee di varie risorse. Le istantanee forniscono copie di point-in-time backup di tali risorse. Esistono molti AWS servizi in grado di supportarti nelle operazioni di backup e ripristino. Per informazioni dettagliate su questi servizi e approcci per il backup

e il ripristino, consulta la guida prescrittiva per il backup e il ripristino e Use backups to recover from security incidents.

Soprattutto in situazioni come un attacco ransomware, è fondamentale che i backup siano ben protetti. Per indicazioni sulla protezione dei backup, consulta Top 10 security best practices for securing backups in AWS. Oltre a proteggere i backup, è necessario sottoporli regolarmente a processi di backup e ripristino per verificare che tecnologia e procedure in uso funzionino come previsto.

Automazione delle funzionalità forensi

Durante un evento di sicurezza, il tuo team di risposta agli incidenti deve essere in grado di raccogliere e analizzare rapidamente le prove, mantenendo al contempo l'accuratezza per il periodo di tempo che circonda l'evento (ad esempio acquisendo i log relativi a un evento o una risorsa specifici o raccogliendo il dump della memoria di un'istanza Amazon). EC2 Per il team addetto a rispondere agli incidenti è difficile e dispendioso in termini di tempo raccogliere manualmente le prove pertinenti, soprattutto se istanze e account sono numerosi. Inoltre, la raccolta manuale può essere soggetta all'errore umano. Per questi motivi, occorre sviluppare e implementare il più possibile l'automazione per le funzionalità forensi.

AWS offre una serie di risorse di automazione per l'analisi forense, elencate nella seguente sezione Risorse. Queste risorse sono esempi di modelli di funzionalità forensi che abbiamo sviluppato, implementate dai clienti Sebbene costituiscano un'utile architettura di riferimento per iniziare, prendi in considerazione la possibilità di modificarli o creare nuovi modelli di automazione per le funzionalità forensi in base ad ambiente, requisiti, strumenti e processi forensi.

Risorse

Documenti correlati:

- AWS Guida alla risposta agli incidenti di sicurezza: sviluppo di funzionalità forensi
- AWS Guida alla risposta agli incidenti di sicurezza - Risorse forensi
- Strategie ambientali di indagine forense nel Cloud AWS
- Come automatizzare la raccolta forense di dischi in AWS
- AWS Guida prescrittiva: automatizza la risposta agli incidenti e l'analisi forense

Video correlati:

- Automating Incident Response and Forensics

Esempi correlati:

- Automated Incident Response and Forensics Framework
- Automated Forensics Orchestrator per Amazon EC2

SEC10-BP04 Sviluppare e testare i playbook di risposta agli incidenti di sicurezza

Una parte fondamentale della preparazione dei processi di risposta agli incidenti è costituita dalla predisposizione di playbook. I playbook di risposta agli incidenti forniscono una serie di indicazioni prescrittive e di passaggi da seguire in caso di evento di sicurezza. Una struttura e passaggi chiari semplificano la risposta e riducono la probabilità di errore umano.

Livello di rischio associato se questa best practice non fosse adottata: medio

Guida all'implementazione

È necessario creare i playbook per scenari di incidenti come:

- Incidenti previsti: i playbook devono essere creati per gli incidenti previsti, tra cui minacce come Denial of Service (DoS), ransomware e la compromissione delle credenziali.
- Rilevamenti o avvisi di sicurezza noti: è necessario creare dei playbook per i risultati e gli avvisi di sicurezza noti, ad esempio i risultati. GuardDuty Potresti ricevere una GuardDuty scoperta e pensare: «E adesso?» Per evitare la cattiva gestione o l'ignoranza di una GuardDuty scoperta, crea un manuale per ogni potenziale scoperta. GuardDuty Alcuni dettagli e linee guida sulla correzione sono disponibili nella documentazione. GuardDuty Vale la pena notare che non GuardDuty è abilitato di default e comporta un costo. Per maggiori dettagli GuardDuty, consulta l'Appendice A: Definizioni delle funzionalità cloud - Visibilità e avvisi.

I playbook devono contenere i passaggi tecnici che un analista della sicurezza deve seguire per indagare e rispondere in modo adeguato a un potenziale incidente di sicurezza.

Passaggi dell'implementazione

Gli elementi da includere in un playbook sono:

- Panoramica del playbook: quale scenario di rischio o incidente affronta questo playbook? Qual è l'obiettivo del playbook?

- Prerequisiti: quali log, meccanismi di rilevamento e strumenti automatizzati sono necessari per questo scenario di incidente? Qual è la notifica prevista?
- Informazioni su comunicazione ed escalation: chi è coinvolto e quali sono le sue informazioni di contatto? Quali sono le responsabilità di ciascuna parte interessata?
- Passaggi di risposta: in tutti i passaggi per la risposta agli incidenti, quali misure tattiche devono essere prese? Quali query deve eseguire l'analista? Quale codice va eseguito per ottenere il risultato desiderato?
  - Individuazione: come verrà individuato l'incidente?
  - Analisi: come verrà determinato l'ambito dell'impatto?
  - Contenimento: come verrà isolato l'incidente per limitarne la portata?
  - Sradicamento: come verrà rimossa la minaccia dall'ambiente?
  - Ripristino: in che modo il sistema o la risorsa interessati verranno riportati in produzione?
- Risultati previsto: dopo l'esecuzione delle query e del codice, qual è il risultato previsto del playbook?

Risorse

Best practice Well-Architected correlate:

- SEC10-BP02 - Sviluppa piani di gestione degli incidenti

Documenti correlati:

- Framework for Incident Response Playbooks
- Develop your own Incident Response Playbooks
- Incident Response Playbook Samples
- Creazione di un runbook di risposta AWS agli incidenti utilizzando i playbook di Jupyter e Lake CloudTrail

SEC10-BP05 Accesso preliminare alla fornitura

Verifica che i soccorritori dispongano in anticipo dell'accesso corretto AWS per ridurre il tempo necessario dalle indagini fino al ripristino.

Anti-pattern comuni:

- Utilizzo dell'account root per la risposta agli incidenti.
- Modifica degli account utente esistenti.
- Manipolazione diretta delle autorizzazioni quando si fornisce l'IAMelevazione dei privilegi. just-in-time

Livello di rischio associato se questa best practice non fosse adottata: medio

Guida all'implementazione

AWS raccomanda di ridurre o eliminare la dipendenza da credenziali di lunga durata laddove possibile, a favore di credenziali temporanee e meccanismi di escalation dei privilegi. just-in-time Le credenziali di lunga durata sono soggette a rischi per la sicurezza e aumentano il sovraccarico operativo. Per la maggior parte delle attività di gestione, nonché per quelle di risposta agli incidenti, si consiglia di implementare la federazione delle identità insieme all'escalation temporanea per l'accesso amministrativo. In questo modello, un utente richiede l'elevazione a un livello di privilegio superiore (come un ruolo di risposta agli incidenti) e, se è idoneo all'elevazione, la richiesta viene inviata al responsabile dell'approvazione. In caso di approvazione della richiesta, l'utente riceve una serie di credenziali AWS temporanee, utilizzabili per completare le proprie attività. Alla scadenza di tali credenziali, l'utente deve inviare una nuova richiesta di elevazione.

Si consiglia l'uso dell'escalation temporanea dei privilegi nella maggior parte degli scenari di risposta agli incidenti. Il modo corretto per eseguire questa operazione prevede l'utilizzo di AWS Security Token Service e policy di sessione per definire l'ambito dell'accesso.

Esistono scenari in cui le identità federate non sono disponibili, come nei seguenti casi:

- Interruzione correlata a un gestore dell'identità digitale (IdP) compromesso.
- Configurazione errata o errore umano che causa l'interruzione del sistema di gestione dell'accesso federato.
- Attività dannose come un evento Distributed Denial of Service (DDoS) o l'indisponibilità del sistema.

Nei casi precedenti, occorre configurare l'accesso di emergenza break glass in modo da consentire l'indagine e la risoluzione tempestiva degli incidenti. Si consiglia di utilizzare un utente, un gruppo o un ruolo con le autorizzazioni appropriate per eseguire attività e accedere alle risorse. AWS Ricorri all'utente root solo per le attività che richiedono le credenziali dell'utente root. Per verificare che chi risponde agli incidenti disponga del livello di accesso corretto AWS e ad altri sistemi pertinenti,

consigliamo di predisporre account dedicati. Gli account richiedono l'accesso con privilegi e devono essere rigorosamente controllati e monitorati. Gli account vanno creati con il minor numero di privilegi richiesti per eseguire le attività e il livello di accesso deve essere basato sui playbook inclusi nel piano di gestione degli incidenti.

Ricorri a utenti e ruoli specifici e dedicati come best practice. L'aumento temporaneo dell'accesso di utenti o ruoli attraverso l'aggiunta di IAM policy rende poco chiaro quale accesso avessero gli utenti durante l'incidente e rischia di non revocare i privilegi aumentati.

È importante rimuovere il maggior numero possibile di dipendenze per verificare che sia possibile ottenere l'accesso nel maggior numero possibile di scenari di errore. A tal fine, create un playbook per verificare che gli utenti che rispondono agli incidenti vengano creati come utenti in un account di sicurezza dedicato e non siano gestiti tramite alcuna soluzione Federation o Single Sign-on () esistente. SSO Ogni singola persona che interviene dopo un incidente deve avere il proprio account denominato. La configurazione dell'account deve applicare una politica di password avanzata e un'autenticazione a più fattori (). MFA Se i playbook di risposta agli incidenti richiedono solo l'accesso a AWS Management Console, l'utente non dovrebbe avere le chiavi di accesso configurate e dovrebbe essere esplicitamente impedito di creare chiavi di accesso. Questo può essere configurato con IAM policy o policy di controllo del servizio (SCPs) come indicato nelle AWS Security Best Practices for. AWS Organizations SCPs Gli utenti non devono disporre di privilegi oltre alla capacità di assumere i ruoli di risposta agli incidenti in altri account.

Durante un incidente, potrebbe essere necessario concedere l'accesso ad altre persone interne o esterne per supportare le attività di analisi, correzione o ripristino. In questo caso, utilizza il meccanismo del playbook menzionato in precedenza e un processo per verificare la revoca immediata di qualsiasi accesso aggiuntivo immediatamente dopo la risoluzione dell'incidente.

Per verificare che l'uso dei ruoli di risposta agli incidenti possa essere monitorato e verificato correttamente, è essenziale che gli IAM account creati a tale scopo non siano condivisi tra individui e che non vengano utilizzati a meno che non siano necessari per un'attività specifica. Utente root dell'account AWS Se è richiesto l'utente root (ad esempio, IAM l'accesso a un account specifico non è disponibile), utilizzate un processo separato con un playbook disponibile per verificare la disponibilità delle credenziali e del token di accesso dell'utente root. MFA

Per configurare IAM le politiche per i ruoli di risposta agli incidenti, prendi in considerazione l'utilizzo di IAMAccess Analyzer per generare politiche basate sui log. AWS CloudTrail In questo caso, concedi l'accesso come amministratore al ruolo di risposta agli incidenti per un account non di produzione e segui i playbook. Al termine, potrà essere creata una policy che consenta solo le azioni da intraprendere. Questa policy potrà quindi essere applicata a tutti i ruoli di risposta agli incidenti

in tutti gli account. Potresti voler creare una IAM policy separata per ogni playbook per consentire una gestione e un controllo più semplici. Esempi di playbook possono essere piani di risposta per ransomware, violazioni dei dati, perdita dell'accesso alla produzione e altri scenari.

Utilizza gli account di risposta agli incidenti per assumere IAMruoli dedicati alla risposta agli incidenti in altri. Account AWS Questi ruoli devono essere configurati in modo che possano essere assunti solo dagli utenti dell'account di sicurezza e la relazione di fiducia deve richiedere che il principale chiamante si sia autenticato utilizzando. MFA I ruoli devono utilizzare politiche ristrette per controllare l'accesso. IAM Assicurati che tutte le `AssumeRole` richieste per questi ruoli siano registrate CloudTrail e avvisate e che tutte le azioni intraprese utilizzando questi ruoli vengano registrate.

Si raccomanda vivamente che sia gli IAM account che i IAM ruoli abbiano nomi chiari per consentirne la facile individuazione nei log. CloudTrail Un esempio potrebbe essere quello di assegnare un nome agli IAM account *<USER_ID>*`-BREAK-GLASS` e ai IAM ruoli`BREAK-GLASS-ROLE`.

CloudTrailviene utilizzato per registrare le API attività negli AWS account e deve essere utilizzato per configurare gli avvisi sull'utilizzo dei ruoli di risposta agli incidenti. Fai riferimento al post del blog sulla configurazione degli avvisi quando vengono utilizzate le chiavi root. Le istruzioni possono essere modificate per configurare la CloudWatch metrica Amazon filter-to-filter sugli `AssumeRole` eventi relativi al IAM ruolo di risposta agli incidenti:

```
{ $.eventName = "AssumeRole" && $.requestParameters.roleArn =
 "<INCIDENT_RESPONSE_ROLE_ARN>" && $.userIdentity.invokedBy NOT EXISTS && $.eventType !
= "AwsServiceEvent" }
```

Vista la probabilità che i ruoli di risposta agli incidenti abbiano un livello di accesso elevato, è importante che questi avvisi vengano inviati a un gruppo ampio e gestiti tempestivamente.

Durante un incidente, è possibile che un soccorritore richieda l'accesso a sistemi che non sono protetti direttamente da. IAM Queste potrebbero includere istanze Amazon Elastic Compute Cloud, database Amazon Relational Database Service o piattaforme (software-as-a-serviceSaaS).
Si consiglia vivamente di utilizzare, anziché utilizzare protocolli nativi come SSH orRDP, AWS Systems Manager Session Managerper tutti gli accessi amministrativi alle EC2 istanze di Amazon. Questo accesso può essere controllato utilizzandoIAM, che è sicuro e verificato. È inoltre possibile automatizzare parti dei playbook mediante i documenti AWS Systems Manager Run Command, in modo da ridurre gli errori degli utenti e migliorare i tempi di ripristino. Per l'accesso ai database e agli strumenti di terze parti, consigliamo di archiviare le credenziali di accesso AWS Secrets Manager e di concedere l'accesso ai ruoli di soccorritore agli incidenti.

Infine, la gestione degli IAM account di risposta agli incidenti deve essere aggiunta ai processi Joiners, Movers e Leavers e rivista e testata periodicamente per verificare che sia consentito solo l'accesso previsto.

Risorse

Documenti correlati:

- Gestione dell'accesso temporaneo elevato all'ambiente AWS
- AWS Guida alla risposta agli incidenti di sicurezza
- AWS Elastic Disaster Recovery
- AWS Systems Manager Incident Manager
- Impostazione di una politica di password dell'account per IAM gli utenti
- Utilizzo dell'autenticazione a più fattori (MFA) in AWS
- Configurazione dell'accesso tra account con MFA
- Utilizzo di IAM Access Analyzer per generare politiche IAM
- Best practice per le politiche AWS Organizations di controllo dei servizi in un ambiente con più account
- Come ricevere notifiche quando vengono utilizzate le chiavi di accesso root dell' AWS account
- Crea autorizzazioni di sessione dettagliate utilizzando politiche gestite IAM

Video correlati:

- Automatizzazione della risposta agli incidenti e delle analisi forensi in AWS
- DIYguida ai runbook, ai report sugli incidenti e alla risposta agli incidenti
- Preparati e rispondi agli incidenti di sicurezza nel tuo ambiente AWS

Esempi correlati:

- Lab: configurazione AWS dell'account e utente root
- Lab: risposta agli incidenti con AWS console e CLI

SEC10-BP06 Strumenti di pre-installazione

Verifica che il team addetto alla sicurezza disponga degli strumenti giusti pre-implementati per ridurre i tempi di indagine fino al ripristino.

Livello di rischio associato se questa best practice non fosse adottata: medio

Guida all'implementazione

Per automatizzare la risposta di sicurezza e le funzioni operative, è possibile utilizzare un set completo di strumenti di. APIs AWS Puoi automatizzare completamente le funzionalità di gestione delle identità, sicurezza della rete, protezione dei dati e monitoraggio e distribuirle utilizzando metodi di sviluppo software comuni già esistenti. Quando crei l'automazione della sicurezza, il sistema può monitorare, rivedere e avviare una risposta, anziché far sì che le persone monitorino la tua posizione di sicurezza e reagiscano manualmente agli eventi.

Se i team di risposta agli incidenti continuano a rispondere agli avvisi nello stesso modo, rischiano il cosiddetto affaticamento dagli avvisi ("alert fatigue"). Ciò significa che, nel corso del tempo, il team può diventare desensibilizzato agli avvisi e commettere errori nella gestione di situazioni ordinarie o farsi sfuggire avvisi insoliti. L'automazione aiuta a evitare l'affaticamento dagli avvisi mediante funzioni che elaborano gli avvisi ripetitivi e ordinari, lasciando alle persone la gestione degli incidenti sensibili e univoci. L'integrazione di sistemi di rilevamento delle anomalie, come Amazon GuardDuty, AWS CloudTrail Insights e Amazon CloudWatch Anomaly Detection, può ridurre l'onere degli avvisi comuni basati su soglie.

Puoi migliorare i processi manuali automatizzando le fasi del processo a livello di programmazione. Dopo aver definito il modello di correzione di un evento, puoi scomporlo in una logica fruibile e scrivere il codice per eseguirla. Il team addetto alla risposta può quindi eseguire il codice per risolvere il problema. Nel corso del tempo, puoi automatizzare più fasi e, infine, gestire automaticamente intere classi di incidenti comuni.

Durante un'indagine di sicurezza, devi essere in grado di esaminare i log pertinenti per registrare e comprendere l'intera portata e la tempistica dell'incidente. I log servono anche per la generazione di avvisi, che indicano il verificarsi di determinate azioni di interesse. È fondamentale selezionare, attivare, memorizzare e impostare i meccanismi di query e recupero e impostare gli avvisi. Inoltre, una soluzione efficace per fornire gli strumenti di ricerca nei dati di log è Amazon Detective.

AWS offre oltre 200 servizi cloud e migliaia di funzionalità. Ti consigliamo di esaminare i servizi in grado di supportare e semplificare la tua strategia di risposta agli incidenti.

Oltre ai log, è necessario sviluppare e implementare una strategia di assegnazione tag. L'etichettatura può aiutare a fornire un contesto relativo allo scopo di una risorsa. AWS e può essere utilizzata anche per l'automazione.

Passaggi dell'implementazione

Seleziona e configura i log per analisi e avvisi

Consulta la seguente documentazione sulla configurazione dei log per la risposta agli incidenti:

- Logging strategies for security incident response
- SEC04-BP01 Configurare la registrazione dei servizi e delle applicazioni

Enable security services to support detection and response

AWS fornisce funzionalità native di investigazione, prevenzione e risposta e altri servizi possono essere utilizzati per progettare soluzioni di sicurezza personalizzate. Per un elenco dei servizi più pertinenti per la risposta agli incidenti di sicurezza, consulta Definizioni delle capacità del cloud.

Sviluppa e implementa una strategia di assegnazione tag

Ottenere informazioni contestuali sul caso d'uso aziendale e sugli stakeholder interni pertinenti che circondano una AWS risorsa può essere difficile. Un modo per farlo è utilizzare i tag, che assegnano metadati alle AWS risorse e sono costituiti da una chiave e da un valore definiti dall'utente. Puoi creare i tag per classificare le risorse per scopo, proprietario, ambiente, tipo di dati elaborati e altri criteri di tua scelta.

Una strategia di tagging coerente può velocizzare i tempi di risposta e ridurre al minimo il tempo dedicato al contesto organizzativo, poiché consente di identificare e distinguere rapidamente le informazioni contestuali su una risorsa. AWS I tag possono anche fungere da meccanismo per avviare le automazioni di risposta. Per maggiori dettagli su cosa taggare, consulta Etichettare le risorse. AWS Dovrai prima definire i tag nella tua organizzazione e quindi implementare e applicare questa strategia di tag. Per maggiori dettagli sull'implementazione e l'applicazione, consulta Implementazione della strategia di etichettatura AWS delle risorse utilizzando le politiche di AWS tag e le politiche di controllo dei servizi () SCPs.

Risorse

Best practice Well-Architected correlate:

- SEC04-BP01 Configurare la registrazione dei servizi e delle applicazioni

- SEC04-BP02 Acquisizione di log, risultati e metriche in posizioni standardizzate

Documenti correlati:

- Logging strategies for security incident response
- Incident response cloud capability definitions

Esempi correlati:

- Rilevamento e risposta alle minacce con Amazon GuardDuty e Amazon Detective
- Workshop Security Hub
- Gestione delle vulnerabilità con Amazon Inspector

SEC10-BP07 Esegui simulazioni

Man mano che le organizzazioni crescono e si evolvono nel tempo, aumentano anche le tipologie di minacce. Per questo motivo, è importante rivedere continuamente le capacità di risposta agli incidenti. L'esecuzione di simulazioni (note anche come giornate di gioco) è un metodo che può essere utilizzato per eseguire questa valutazione. Le simulazioni utilizzano scenari di eventi di sicurezza reali progettati per imitare le tattiche, le tecniche e le procedure di un autore della minaccia (TTPs) e consentire a un'organizzazione di esercitare e valutare le proprie capacità di risposta agli incidenti rispondendo a questi finti eventi informatici così come potrebbero verificarsi nella realtà.

Vantaggi dell'adozione di questa best practice: le simulazioni offrono una serie di vantaggi.

- Convalida della preparazione informatica e sviluppo della fiducia dei team di risposta agli incidenti.
- Verifica della precisione e dell'efficienza di strumenti e flussi di lavoro.
- Perfezionamento dei metodi di comunicazione ed escalation in linea con il piano di risposta agli incidenti.
- Opportunità di rispondere per i vettori meno comuni.

Livello di rischio associato se questa best practice non fosse adottata: medio

Guida all'implementazione

Esistono tre tipi principali di simulazioni:

- Simulazioni di situazioni di emergenza le simulazioni di situazioni di emergenza sono sessioni basate sulla discussione che coinvolgono le varie parti interessate alla risposta agli incidenti per mettere in pratica ruoli e responsabilità e utilizzare strumenti e playbook di comunicazione consolidati. Lo svolgimento dell'esercitazione può in genere essere eseguito in un'intera giornata in un luogo virtuale, in un luogo fisico o in una combinazione di questi tipi di luogo. Poiché è basato sulla discussione, questo tipo di esercitazione si concentra su processi, persone e collaborazione. La tecnologia è parte integrante della discussione, ma l'uso effettivo di strumenti o script di risposta agli incidenti in genere non rientra in questo tipo di simulazione.
- Esercitazioni con il team viola: questo tipo di esercitazioni aumenta il livello di collaborazione tra i team di risposta agli incidenti (team blu) e gli attori delle minacce simulate (team rosso). Il team blu è composto da membri del centro operativo di sicurezza (SOC), ma può includere anche altre parti interessate che potrebbero essere coinvolte durante un vero evento informatico. Il team rosso è composto da un team responsabile dei test di penetrazione o da parti interessate chiave esperte in materia di sicurezza informatica. Il team rosso lavora assieme ai coordinatori dell'esercitazione durante la progettazione di uno scenario in modo che questi sia accurato e fattibile. Durante le esercitazioni di squadra in viola, l'attenzione principale è rivolta ai meccanismi di rilevamento, agli strumenti e alle procedure operative standard (SOPs) a supporto degli sforzi di risposta agli incidenti.
- Esercitazioni con il team rosso: durante un'esercitazione con il team rosso, l'attacco (team rosso) effettua una simulazione per raggiungere un determinato obiettivo o una serie di obiettivi da un ambito predeterminato. I difensori (team blu) non saranno necessariamente a conoscenza della portata e della durata dell'esercitazione, il che fornisce una valutazione più realistica di come risponderebbero a un incidente reale. Poiché le esercitazioni con il team rosso possono basarsi su test invasivi, procedi con cautela e implementa controlli per verificare che l'esercitazione non causi danni effettivi all'ambiente.

Prendi in considerazione la possibilità di svolgere simulazioni informatiche a intervalli regolari. Ogni tipo di esercitazione può offrire vantaggi unici ai partecipanti e all'organizzazione nel suo insieme; potresti, quindi, scegliere di iniziare con tipi di simulazione meno complessi (come le simulazioni di situazioni di emergenza) e passare a tipi di simulazione più complessi (esercitazioni del team rosso). È necessario selezionare un tipo di simulazione in base alla maturità, alle risorse e ai risultati desiderati a livello di sicurezza. Alcuni clienti potrebbero scegliere di non eseguire le esercitazioni del team rosso a causa della loro complessità e dei loro costi.

Passaggi dell'implementazione

Indipendentemente dal tipo di simulazione scelto, le simulazioni sono in genere caratterizzate dai seguenti passaggi di implementazione:

1. Definisci gli elementi principali dell'esercitazione: definisci scenario e obiettivi della simulazione. Lo scenario e gli obiettivi dovrebbero essere entrambi accettati dalla leadership.
2. Identifica le parti interessate principali: come minimo, un'esercitazione prevede la presenza di coordinatori e partecipanti. A seconda dello scenario, potrebbero essere coinvolte altre parti interessate come la leadership legale, delle comunicazioni o esecutiva.
3. Crea ed esegui il test dello scenario: potrebbe essere necessario ridefinire lo scenario durante la creazione se risulta impossibile implementare elementi specifici. Come risultato di questa fase è previsto uno scenario definitivo.
4. Fai svolgere la simulazione: il tipo di simulazione determina il tipo di svolgimento usato (uno scenario basato su supporto cartaceo o uno scenario con simulazione altamente tecnologica). I coordinatori dovrebbero allineare le loro tattiche di svolgimento agli oggetti dell'esercitazione e dovrebbero coinvolgere tutti i partecipanti ove possibile per ottimizzare i benefici.
5. Sviluppa il rapporto post-azione (AAR): identifica le aree che sono andate bene, quelle che possono essere migliorate e le potenziali lacune. AARDovrebbero misurare l'efficacia della simulazione e la risposta del team all'evento simulato in modo da poter monitorare i progressi nel tempo con simulazioni future.

Risorse

Documenti correlati:

- AWS Guida alla risposta agli incidenti

Video correlati:

- AWS GameDay - Edizione di sicurezza

SEC10-BP08 Stabilire un framework per imparare dagli incidenti

L'implementazione di un framework basato sulle lezioni apprese e di una capacità di analisi delle cause principali non solo contribuisce a migliorare le capacità di risposta agli incidenti, ma aiuta anche a prevenire il ripetersi dell'incidente. Imparando da ogni incidente, puoi evitare di ripetere gli

errori, i rischi o le configurazioni non valide, non solo migliorando il tuo livello di sicurezza, ma anche riducendo al minimo il tempo speso in situazioni evitabili.

Livello di rischio associato se questa best practice non fosse adottata: medio

Guida all'implementazione

È importante implementare un framework basato sulle lezioni apprese in grado di stabilire e raggiungere, a un livello elevato, i seguenti punti:

- Quando si tiene un framework basato sulle lezioni apprese?
- Cosa comporta il processo basato sulle lezioni apprese?
- Come viene eseguito un framework basato sulle lezioni apprese?
- Chi è coinvolto nel processo e in che modo?
- Come vengono identificate le aree di miglioramento?
- In che modo garantisci che i miglioramenti vengano monitorati e implementati in modo efficace?

Il framework non deve concentrarsi sugli individui, ma sul miglioramento di strumenti e processi.

Passaggi dell'implementazione

A parte i risultati di alto livello sopra elencati, è importante porsi le domande giuste per trarre il massimo valore (informazioni che portano a miglioramenti attuabili) dal processo. Considera queste domande per iniziare a promuovere le discussioni sulle lezioni apprese:

- Qual è stato l'incidente?
- Quando è stato identificato per la prima volta l'incidente?
- Come è stato identificato?
- Quali sistemi hanno avvisato dell'attività?
- Quali sistemi, servizi e dati sono stati coinvolti?
- Cosa è successo nello specifico?
- Cosa ha funzionato bene?
- Cosa non ha funzionato bene?
- Quale processo o quali procedure non sono riusciti a scalare per rispondere all'incidente?
- Cosa può essere migliorato nelle seguenti aree:
  - Persone

    - Le persone da contattare erano effettivamente disponibili e l'elenco dei contatti era aggiornato?
    - Le persone presentavano lacune nella formazione o nelle capacità necessarie per rispondere e indagare efficacemente sull'incidente?
    - Le risorse appropriate erano pronte e disponibili?
- Processo
    - Sono stati seguiti i processi e le procedure?
    - I processi e le procedure erano documentati e disponibili per questo tipo di incidente?
    - Mancavano i processi e le procedure richiesti?
    - Il team di risposta è stato in grado di accedere tempestivamente alle informazioni necessarie per rispondere al problema?
- Tecnologia
    - I sistemi di avviso esistenti hanno identificato e segnalato efficacemente l'attività?
    - Come avremmo potuto ridurlo del 50%? time-to-detection
    - Gli avvisi esistenti devono essere migliorati o è necessario creare nuovi avvisi per questo (tipo di) incidente?
    - Gli strumenti esistenti hanno consentito un'indagine efficace (ricerca/analisi) dell'incidente?
    - Cosa si può fare per identificare prima questo tipo di incidente?
    - Cosa si può fare per evitare che questo tipo di incidente si ripeta?
    - A chi appartiene il piano di miglioramento e come verifichi che sia stato implementato?
    - Qual è la tempistica per l'implementazione e il test del monitoraggio aggiuntivo o dei controlli e dei processi preventivi?

Questo elenco non è esaustivo, ma può fungere da punto di partenza per individuare quali sono le esigenze dell'organizzazione e dell'attività e come analizzarle per imparare in modo più efficace dagli incidenti e migliorare costantemente il proprio livello di sicurezza. La cosa più importante è iniziare incorporando le lezioni apprese come parte standard del processo di risposta agli incidenti, della documentazione e delle aspettative di tutti le parti interessate.

Risorse

Documenti correlati:

- AWS Security Incident Response Guide - Establish a framework for learning from incidents

- NCSCCAForientamento - Lezioni apprese

Sicurezza delle applicazioni

Domanda

- SEC11. Come si incorporano e convalidano le proprietà di sicurezza delle applicazioni nell'intero ciclo di vita di progettazione, sviluppo e implementazione?

SEC11. Come si incorporano e convalidano le proprietà di sicurezza delle applicazioni nell'intero ciclo di vita di progettazione, sviluppo e implementazione?

La formazione del personale, l'esecuzione di test tramite automazione, l'identificazione delle dipendenze e la convalida delle proprietà di sicurezza di strumenti e applicazioni riducono la probabilità del verificarsi di problemi di sicurezza nei carichi di lavoro di produzione.

Best practice

- SEC11-BP01 Train per la sicurezza delle applicazioni
- SEC11-BP02 Automatizza i test durante tutto il ciclo di vita di sviluppo e rilascio
- SEC11-BP03 Eseguire test di penetrazione regolari
- SEC11-BP04 Revisioni del codice manuale
- SEC11-BP05 Centralizza i servizi per pacchetti e dipendenze
- SEC11-BP06 Implementazione del software a livello di codice
- SEC11-BP07 Valutare regolarmente le proprietà di sicurezza delle condotte
- SEC11-BP08 Crea un programma che incorpori la titolarità della sicurezza nei team addetti ai carichi di lavoro

SEC11-BP01 Train per la sicurezza delle applicazioni

Fornisci formazione sulle procedure comuni agli sviluppatori nell'organizzazione in modo da garantire la sicurezza dello sviluppo e del funzionamento delle applicazioni. L'adozione di procedure di sviluppo incentrate sulla sicurezza riduce la probabilità di riscontrare problemi rilevati solo nella fase di revisione della sicurezza.

Risultato desiderato: il software è progettato e realizzato nell'ottica della sicurezza. Quando gli sviluppatori in un'organizzazione ricevono formazione su procedure di sviluppo sicure che partono

da un modello di rischio, la qualità e la sicurezza complessive del software prodotto sono migliori. Questo approccio può ridurre il tempo necessario per distribuire il software o le funzionalità, in quanto saranno necessarie meno correzioni dopo la fase di revisione della sicurezza.

Ai fini di questa best practice, per sviluppo sicuro si intende il software in fase di scrittura e gli strumenti o i sistemi che supportano il ciclo di vita dello sviluppo del software (). SDLC

Anti-pattern comuni:

- Valutazione delle proprietà di sicurezza di un sistema solo in fase di revisione della sicurezza.
- Assegnazione di tutte le decisioni in materia di sicurezza al team responsabile della sicurezza.
- Mancata comunicazione del SDLC rapporto tra le decisioni prese e le aspettative o le politiche generali di sicurezza dell'organizzazione.
- Svolgimento del processo di revisione della sicurezza in una fase troppo tardiva.

Vantaggi dell'adozione di questa best practice:

- Migliore identificazione dei requisiti aziendali per la sicurezza all'inizio del ciclo di sviluppo.
- Capacità di identificare e correggere più rapidamente possibili problemi di sicurezza, per una distribuzione più rapida delle funzionalità.
- Migliore qualità del software e dei sistemi.

Livello di rischio associato se questa best practice non fosse adottata: medio

Guida all'implementazione

Fornisci formazione agli sviluppatori nell'organizzazione. Partire con un corso sulla modellazione delle minacce è una buona base di partenza nella formazione sulla sicurezza. Idealmente, gli sviluppatori devono poter accedere in modalità self-service a informazioni pertinenti ai propri carichi di lavoro. Questo accesso può aiutarli a prendere decisioni informate sulle proprietà di sicurezza dei sistemi sviluppati senza dover chiedere a un altro team. Il processo di coinvolgimento del team responsabile della sicurezza nelle revisioni deve essere definito chiaramente e facile da seguire. Le fasi del processo di revisione vanno incluse nella formazione sulla sicurezza. Se disponibili, modelli o schemi di implementazione noti devono essere facili da trovare e collegare ai requisiti complessivi per la sicurezza. Prendi in considerazione l'utilizzo di AWS CloudFormation, costrutti di AWS Cloud Development Kit (AWS CDK), Service Catalog o altri strumenti di creazione di modelli per ridurre la necessità di configurazioni personalizzate.

Passaggi dell'implementazione

- Iscrivi gli sviluppatori a un corso sulla modellazione delle minacce per creare una buona base e formarli su come pensare alla sicurezza.
- Fornisci accesso alla AWS Training certificazione, al settore o alla formazione per i AWS partner.
- Prevedi corsi di formazione sul processo di revisione della sicurezza dell'organizzazione, che spieghino la suddivisione delle responsabilità tra il team responsabile della sicurezza, i team del carico di lavoro e altre parti interessate
- Pubblica linee guida self-service su come soddisfare i requisiti di sicurezza, inclusi esempi e modelli di codice, se disponibili.
- Richiedi regolarmente ai team di sviluppo feedback sull'esperienza durante il processo di revisione della sicurezza e la formazione correlata e usalo per migliorare le procedure.
- Usa campagne di simulazione o bug bash per ridurre il numero di problemi e migliorare le competenze degli sviluppatori.

Risorse

Best practice correlate:

- SEC11-BP08 Crea un programma che incorpori la titolarità della sicurezza nei team addetti ai carichi di lavoro

Documenti correlati:

- AWS Training e certificazione
- How to think about cloud security governance
- How to approach threat modeling
- Accelerare la formazione — The AWS Skills Guild

Video correlati:

- Proactive security: Considerations and approaches

Esempi correlati:

- Workshop on threat modeling

- Industry awareness for developers

Servizi correlati:

- AWS CloudFormation
- AWS Cloud Development Kit (AWS CDK) ()AWS CDK Costrutti
- Service Catalog
- AWS BugBust

SEC11-BP02 Automatizza i test durante tutto il ciclo di vita di sviluppo e rilascio

Automatizza i test per le proprietà di sicurezza lungo il ciclo di vita di sviluppo e test. L'automazione semplifica l'identificazione coerente e ripetibile dei potenziali problemi nel software prima del rilascio, riducendo il rischio di riscontrare problemi di sicurezza nel software fornito.

Risultato desiderato: l'obiettivo dei test automatizzati è fornire una soluzione programmatica per l'individuazione di potenziali problemi nelle fasi iniziali e spesso durante l'intero ciclo di vita dello sviluppo. Automatizzando i test di regressione, puoi ripetere l'esecuzione di test funzionali e non funzionali per verificare che il software testato in precedenza continui ad avere le prestazioni previste dopo una modifica. Quando definisci test di unità di sicurezza per verificare la presenza di configurazioni errate comuni, come autorizzazioni non corrette o mancanti, puoi identificare e correggere i problemi all'inizio del processo di sviluppo.

Per l'automazione dei test vengono usati casi di test dedicati per la convalida delle applicazioni, in base ai requisiti e alle funzionalità desiderate. Il risultato dei test automatici è basato sul confronto dell'output del test generato con quello previsto, che accelera l'intero ciclo di vita dei test. Metodologie di test come i test di regressione e le suite di test di unità sono ideali per l'automazione. L'automazione dei test delle proprietà di sicurezza permette agli sviluppatori di ricevere in automatico feedback senza attendere una revisione della sicurezza. I test automatici sotto forma di analisi statica o dinamica del codice possono migliorare la qualità del codice e semplificare il rilevamento dei potenziali problemi software all'inizio del ciclo di vita di sviluppo.

Anti-pattern comuni:

- Mancata comunicazione dei casi di test e dei risultati dei test automatici.
- Esecuzione dei test solo immediatamente prima di un rilascio.
- Automazione dei casi di test con requisiti che cambiano spesso.

- Assenza di linee guida su come gestire i risultati dei test di sicurezza.

Vantaggi dell'adozione di questa best practice:

- Riduzione della dipendenza da valutazioni personali delle proprietà di sicurezza dei sistemi.
- Migliore coerenza grazie a esiti uniformi tra più flussi di lavoro.
- Minore probabilità di introdurre problemi di sicurezza nel software di produzione.
- Intervallo di tempo più breve tra l'individuazione e la correzione grazie all'identificazione più tempestiva dei problemi software.
- Maggiore visibilità su comportamenti sistematici o ripetuti tra più flussi di lavoro, utile per favorire miglioramenti in tutta l'organizzazione.

Livello di rischio associato se questa best practice non fosse adottata: medio

Guida all'implementazione

Durante lo sviluppo del software, adotta diversi meccanismi di test in modo da avere la certezza di testare l'applicazione per requisiti funzionali, basati sulla logica di business, e non funzionali, incentrati sull'affidabilità, sulle prestazioni e sulla sicurezza dell'applicazione.

Static Application Security Testing (SAST) analizza il codice sorgente alla ricerca di modelli di sicurezza anomali e fornisce indicazioni sul codice soggetto a difetti. SASTsi basa su input statici, come la documentazione (specifiche dei requisiti, documentazione di progettazione e specifiche di progettazione) e il codice sorgente dell'applicazione per verificare la presenza di una serie di problemi di sicurezza noti. Gli analizzatori di codice statici possono contribuire ad accelerare l'analisi di volumi elevati di codice. Il NISTQuality Group fornisce un confronto tra Source Code Security Analyzers, che include strumenti open source per scanner di codici a byte e scanner di codici binari.

Completa i test statici con metodologie di analisi dinamica (security testing (DAST), che eseguono test sull'applicazione in esecuzione per identificare comportamenti potenzialmente imprevisti. I test dinamici consentono di individuare potenziali problemi non rilevabili tramite l'analisi statica. L'esecuzione di test nelle fasi di repository, compilazione e pipeline del codice permette di verificare potenziali problemi di tipi diversi, evitandone la presenza nel codice. Amazon CodeWhisperer fornisce consigli sul codice, inclusa la scansione di sicurezza, nel builder. IDE Amazon CodeGuru Reviewer è in grado di identificare problemi critici, problemi di sicurezza e hard-to-find bug durante lo sviluppo di applicazioni e fornisce consigli per migliorare la qualità del codice.

Il workshop Security for AWS Developers utilizza strumenti per sviluppatori, come, and AWS CodeBuildAWS CodeCommitAWS CodePipeline, per l'automazione della pipeline di rilascio che include SAST e DAST collauda metodologie.

Man mano che procediSDLC, stabilisci un processo iterativo che includa revisioni periodiche delle applicazioni con il tuo team di sicurezza. Il feedback raccolto da queste revisioni della sicurezza deve essere affrontato e convalidato come parte della revisione dell'idoneità per il rilascio. Queste revisioni permettono di stabilire una solida posizione di sicurezza per l'applicazione e forniscono agli sviluppatori feedback di utilità pratica per affrontare i potenziali problemi.

Passaggi dell'implementazione

- Implementa strumenti CI/CD coerentiIDE, di revisione del codice e strumenti CI/CD che includano test di sicurezza.
- Valuta dove è SDLC opportuno bloccare le pipeline invece di limitarsi a notificare ai costruttori che i problemi devono essere risolti.
- Il workshop Security for Developers offre un esempio di integrazione di test statici e dinamici all'interno di una pipeline di rilascio.
- L'esecuzione di test o analisi del codice utilizzando strumenti automatizzati, come Amazon CodeWhisperer integrato con lo sviluppatore IDEs e Amazon CodeGuru Reviewer per la scansione del codice su commit, aiuta i costruttori a ottenere feedback al momento giusto.
- Quando crei utilizzando AWS Lambda, puoi usare Amazon Inspector per scansionare il codice dell'applicazione nelle tue funzioni.
- Se le pipeline CI/CD includono test automatici, devi usare un sistema di gestione dei ticket per tenere traccia della notifica e della correzione dei problemi software.
- Per test di sicurezza che possono generare esiti, il collegamento a linee guida per la correzione permette agli sviluppatori di migliorare la qualità del codice.
- Analizza regolarmente gli esiti ottenuti dagli strumenti automatici per definire le priorità delle successive iniziative di automazione, formazione degli sviluppatori o creazione di campagne di sensibilizzazione.

Risorse

Documenti correlati:

- Distribuzione e implementazione continue

- AWS DevOps Partner di competenza
- AWS Security Competency Partners per la sicurezza delle applicazioni
- Choosing a Well-Architected CI/CD approach
- Monitoraggio CodeCommit degli eventi in Amazon EventBridge e Amazon CloudWatch Events
- Rilevamento di segreti in Amazon CodeGuru Review
- Accelera le implementazioni AWS con una governance efficace
- How AWS approaches automating safe, hands-off deployments

Video correlati:

- Hands-off: Automating continuous delivery pipelines at Amazon
- Automating cross-account CI/CD pipelines

Esempi correlati:

- Industry awareness for developers
- AWS CodePipeline Governance () GitHub
- Workshop Security for Developers

SEC11-BP03 Eseguire test di penetrazione regolari

Esegui regolarmente test di penetrazione sul software. Questo meccanismo ti consente di identificare potenziali problemi relativi al software che non possono essere rilevati dai test automatizzati o dalla revisione manuale del codice e può anche aiutarti a capire l'efficacia dei tuoi controlli di rilevamento. I test di penetrazione devono determinare se il software può essere reso operativo in modi imprevisti, ad esempio esponendo dati che da proteggere o concedendo autorizzazioni più elevate del previsto.

Risultato desiderato: utilizzo del test di penetrazione per rilevare, correggere e convalidare le proprietà di sicurezza dell'applicazione. I test di penetrazione regolari e programmati devono essere eseguiti come parte del ciclo di vita di sviluppo del software (). SDLC Gli esiti ottenuti dai test di penetrazione devono essere gestiti prima del rilascio del software. Devi analizzare gli esiti dei test di penetrazione per identificare l'eventuale presenza di problemi identificabili con l'automazione. Un processo di esecuzione di test di penetrazione regolare e ripetibile, con un meccanismo di feedback attivo, aiuta a stabilire linee guida per gli sviluppatori e migliora la qualità del software.

Anti-pattern comuni:

- Esecuzione di test di penetrazione solo per problemi di sicurezza noti o comuni.
- Esecuzione di test di penetrazione delle applicazioni senza gli strumenti e le librerie di terze parti dipendenti.
- Esecuzione di test di penetrazione solo per i problemi di sicurezza relativi ai pacchetti, senza valutare la logica di business implementata.

Vantaggi dell'adozione di questa best practice:

- Maggiore certezza riguardo alle proprietà di sicurezza del software prima del rilascio.
- Opportunità di identificare i modelli comportamentali preferiti delle applicazioni, per una migliore qualità del software.
- Presenza di un ciclo di feedback che identifica all'inizio del ciclo di sviluppo i punti in cui l'automazione o una formazione aggiuntiva possono migliorare le proprietà di sicurezza del software.

Livello di rischio associato se questa best practice non fosse adottata: elevato

Guida all'implementazione

I test di penetrazione sono un esercizio strutturato per l'esecuzione di test di sicurezza in cui vengono eseguiti scenari di violazione della sicurezza pianificati per rilevare, correggere e convalidare i controlli di sicurezza. I test di penetrazione partono dalla ricognizione, durante la quale si raccolgono dati in base all'attuale progettazione dell'applicazione e alle sue dipendenze. Viene creato ed eseguito un elenco selezionato di scenari di test specifici per la sicurezza. Lo scopo principale di questi test è rivelare i problemi di sicurezza nell'applicazione che potrebbero essere sfruttati per ottenere l'accesso indesiderato all'ambiente o l'accesso non autorizzato ai dati. Devi eseguire test di penetrazione quando lanci nuove funzionalità o ogni volta che l'applicazione viene sottoposta a modifiche importanti durante l'implementazione tecnica o di funzioni.

Devi identificare la fase più appropriata del ciclo di vita di sviluppo in cui eseguire i test di penetrazione. Questi test devono essere eseguiti nelle fasi finali, in modo che la funzionalità del sistema sia vicina allo stato di rilascio previsto, ma con tempo sufficiente per la correzione di eventuali problemi.

Passaggi dell'implementazione

- Adotta un processo strutturato per definire l'ambito dei test di penetrazione. Basare il processo sul modello di minaccia costituisce una buona soluzione per mantenere il contesto.
- Identifica la fase più appropriata del ciclo di vita di sviluppo in cui eseguire test di penetrazione. Questi devono avvenire quando sono previste modifiche minime nell'applicazione, ma quando vi è ancora tempo sufficiente per apportare eventuali correzioni.
- Prepara gli sviluppatori su cosa aspettarsi dagli esiti dei test di penetrazione e su come ottenere informazioni sulla correzione.
- Usa strumenti per accelerare il processo di esecuzione dei test di penetrazione automatizzando test comuni o ripetibili.
- Analizza gli esiti dei test di penetrazione per identificare problemi di sicurezza sistematici e usa questi dati per definire altri test automatici e formazione continua per gli sviluppatori.

Risorse

Best practice correlate:

- SEC11-BP01 Train per la sicurezza delle applicazioni
- SEC11-BP02 Automatizza i test durante tutto il ciclo di vita di sviluppo e rilascio

Documenti correlati:

- AWS Il test di penetrazione fornisce una guida dettagliata per i test di penetrazione su AWS
- Accelera le implementazioni con una governance efficace AWS
- AWS Security Competency Partners
- Modernizza la tua architettura di test di penetrazione su AWS Fargate
- AWS Simulatore di iniezione di guasti

Esempi correlati:

- Automatizza i API test con AWS CodePipeline () GitHub
- Helper di sicurezza automatizzato () GitHub

SEC11-BP04 Revisioni del codice manuale

Esegui una revisione manuale del codice del software che produci. Questo processo ti consente di assicurarti che chi ha scritto il codice non sia l'unica persona a controllarne la qualità.

Risultato desiderato: l'inclusione di una fase di revisione manuale del codice durante lo sviluppo aumenta la qualità del software scritto, migliora le competenze dei membri meno esperti del team e consente di identificare i luoghi di possibile uso dell'automazione. Le revisioni manuali del codice possono essere supportate da strumenti e test automatici.

Anti-pattern comuni:

- Nessuna revisione del codice prima dell'implementazione.
- Scrittura e revisione del codice effettuate dalla stessa persona.
- Mancato utilizzo dell'automazione per semplificare o orchestrare le revisioni del codice.
- Mancata formazione degli sviluppatori sulla sicurezza dell'applicazione prima di eseguire la revisione del codice.

Vantaggi dell'adozione di questa best practice:

- Migliore qualità del codice.
- Maggiore coerenza dello sviluppo del codice attraverso il riutilizzo di approcci comuni.
- Riduzione del numero di problemi riscontrati durante i test di penetrazione e nelle fasi successive.
- Migliore circolazione delle informazioni all'interno del team.

Livello di rischio associato se questa best practice non fosse adottata: medio

Guida all'implementazione

La fase di revisione deve essere implementata nell'ambito del flusso complessivo di gestione del codice. Le specifiche dipendono dall'approccio usato per la diramazione, le richieste pull e l'unione. Potresti utilizzare soluzioni di terze parti come GitHub GitLab, AWS CodeCommit o Bitbucket. Indipendentemente dal metodo usato, è importante verificare che i processi richiedano la revisione del codice prima dell'implementazione in un ambiente di produzione. L'utilizzo di strumenti come Amazon CodeGuru Reviewer può semplificare l'orchestrazione del processo di revisione del codice.

Passaggi dell'implementazione

- Implementa una fase di revisione manuale nell'ambito del flusso di gestione del codice ed esegui la revisione prima di continuare.
- Prendi in considerazione Amazon CodeGuru Reviewer per la gestione e l'assistenza nelle revisioni del codice.
- Implementa un flusso di approvazione che richieda il completamento di una revisione prima che il codice possa passare alla fase successiva.
- Verifica che sia stato definito un processo per identificare i problemi riscontrati durante le revisioni manuali del codice rilevabili in automatico.
- Integra la fase di revisione manuale del codice in modo che sia allineata alle pratiche di sviluppo del codice.

Risorse

Best practice correlate:

- SEC11-BP02 Automatizza i test durante tutto il ciclo di vita di sviluppo e rilascio

Documenti correlati:

- Lavorare con le richieste pull nei repository AWS CodeCommit
- Utilizzo dei modelli di regole di approvazione in AWS CodeCommit
- Informazioni sulle pull request in GitHub
- Automatizza le revisioni del codice con Amazon Reviewer CodeGuru
- Rilevamento automatico di vulnerabilità e bug di sicurezza nelle pipeline CI/CD con Amazon Reviewer CodeGuru CLI

Video correlati:

- Miglioramento continuo della qualità del codice con Amazon CodeGuru

Esempi correlati:

- Workshop Security for Developers

SEC11-BP05 Centralizza i servizi per pacchetti e dipendenze

Fornisci servizi centralizzati per permettere ai team di sviluppo di ottenere pacchetti software e altre dipendenze. Questo approccio permette la convalida dei pacchetti prima di includerli nel software scritto e fornisce un'origine dati per l'analisi del software usato nell'organizzazione.

Risultato desiderato: oltre al codice scritto, il software si compone di un insieme di altri pacchetti software. Ciò semplifica l'utilizzo di implementazioni di funzionalità che vengono utilizzate ripetutamente, come un JSON parser o una libreria di crittografia. La centralizzazione logica delle origini per questi pacchetti e dipendenze fornisce un meccanismo tramite il quale i team responsabili della sicurezza possono convalidare le proprietà dei pacchetti prima che vengano usati. Questo approccio riduce anche il rischio di un problema imprevisto causato da una modifica in un pacchetto esistente o dall'aggiunta da parte dei team di sviluppo di pacchetti arbitrari direttamente da Internet. Usa questo approccio insieme ai flussi di test manuali e automatici per garantire ulteriormente la qualità del software sviluppato.

Anti-pattern comuni:

- Recupero di pacchetti da repository arbitrari su Internet.
- Mancata esecuzione di test sui nuovi pacchetti prima di renderli disponibili agli sviluppatori.

Vantaggi dell'adozione di questa best practice:

- Migliore comprensione dei pacchetti usati nel software sviluppato.
- Capacità di informare i team responsabili del carico di lavoro quando un pacchetto deve essere aggiornato in base alle informazioni su chi usa cosa.
- Minor rischio di includere nel software un pacchetto con problemi.

Livello di rischio associato se questa best practice non fosse adottata: medio

Guida all'implementazione

Fornisci servizi centralizzati per i pacchetti e le dipendenze in modo da semplificarne l'uso per gli sviluppatori. La centralizzazione dei servizi può essere eseguita in modo logico anziché implementarli come sistema monolitico. Questo approccio permette di fornire servizi in modo da soddisfare le esigenze degli sviluppatori. È necessario implementare un modo efficiente per aggiungere pacchetti al repository quando vengono effettuati aggiornamenti o emergono nuovi requisiti. AWS servizi

come soluzioni AWS partner simili AWS CodeArtifacto simili forniscono un modo per fornire questa funzionalità.

Passaggi dell'implementazione:

- Implementa un servizio di repository centralizzato in modo logico che sia disponibile in tutti gli ambienti in cui viene sviluppato il software.
- Includi l'accesso al repository come parte del processo di provisioning automatico dell' Account AWS .
- Crea automazione per testare i pacchetti prima della loro pubblicazione in un repository.
- Gestisci le metriche dei pacchetti, dei linguaggi e dei team usati più comunemente e con la maggiore quantità di modifiche.
- Offri ai team di sviluppo un meccanismo automatico per richiedere nuovi pacchetti e fornire feedback.
- Analizza regolarmente i pacchetti nel repository per identificare il possibile impatto di nuovi problemi riscontrati.

Risorse

Best practice correlate:

- SEC11-BP02 Automatizza i test durante tutto il ciclo di vita di sviluppo e rilascio

Documenti correlati:

- Accelera le implementazioni AWS con una governance efficace
- Rafforza la sicurezza dei tuoi pacchetti con il toolkit CodeArtifact Package Origin Control
- Rilevamento di problemi di sicurezza nella registrazione con Amazon Reviewer CodeGuru
- Livelli della catena di fornitura per gli artefatti software () SLSA

Video correlati:

- Proactive security: Considerations and approaches
- The AWS Philosophy of Security (re:Invent 2017)
- When security, safety, and urgency all matter: Handling Log4Shell

Esempi correlati:

- Pipeline di pubblicazione di pacchetti multiregione () GitHub
- Pubblicazione dei moduli Node.js sull' AWS CodeArtifact utilizzo AWS CodePipeline di () GitHub
- AWS CDK Esempio di CodeArtifact pipeline Java () GitHub
- Distribuisci in privato. NET NuGet pacchetti con AWS CodeArtifact (GitHub)

SEC11-BP06 Implementazione del software a livello di codice

Esegui implementazioni programmatiche del software laddove possibile. Questo approccio riduce la probabilità che un'implementazione non riesca o che si verifichi un problema imprevisto a causa dell'errore umano.

Risultato desiderato: tenere le persone lontane dai dati è un principio chiave per lo sviluppo sicuro in Cloud AWS. Questo principio include anche il modo in cui viene implementato il software.

I vantaggi legati alla scelta di non affidare a persone l'implementazione del software è la migliore garanzia che la soluzione implementata sia esattamente identica a quella testata e che l'implementazione verrà eseguita in modo coerente ogni volta. Il software non deve essere modificato in modo da funzionare in ambienti diversi. Usando i principi dello sviluppo di applicazioni a dodici fattori, in particolare l'esternalizzazione della configurazione, puoi implementare lo stesso codice in più ambienti senza richiedere modifiche. La firma crittografica dei pacchetti software è un ottimo metodo per verificare che non vengano apportate modifiche tra ambienti. Il risultato complessivo di questo approccio è la riduzione dei rischi nel processo di modifica e il miglioramento della coerenza delle versioni del software.

Anti-pattern comuni:

- Implementazione manuale del software nell'ambiente di produzione.
- Applicazione manuale di modifiche al software per soddisfare i requisiti di ambienti diversi.

Vantaggi dell'adozione di questa best practice:

- Maggiore affidabilità del processo di rilascio del software.
- Riduzione dei rischi legati a modifiche errate che hanno impatto sulla funzionalità aziendale.
- Processi di rilascio più frequenti grazie a un rischio di modifica minimo.
- Funzionalità di rollback automatiche in caso di eventi imprevisti durante l'implementazione.

- Possibilità di usare la crittografia per dimostrare che il software implementato è esattamente identico a quello testato.

Livello di rischio associato se questa best practice non fosse adottata: elevato

Guida all'implementazione

Costruisci la tua Account AWS struttura per rimuovere l'accesso umano persistente dagli ambienti e utilizza gli strumenti CI/CD per eseguire le implementazioni. Progetta le tue applicazioni affinché i dati di configurazione specifici dell'ambiente provengano da una fonte esterna, come AWS Systems Manager Parameter Store. Firma i pacchetti dopo che vengono testati e convalida le firme durante l'implementazione. Configura le pipeline CI/CD per il push del codice delle applicazioni e usa valori canary per confermare la corretta esecuzione dell'implementazione. Utilizza strumenti come AWS CloudFormation o AWS CDK per definire l'infrastruttura, quindi utilizza AWS CodeBuild e AWS CodePipeline per l'esecuzione delle operazioni CI/CD.

Passaggi dell'implementazione

- Crea pipeline CI/CD ben definite per semplificare il processo di implementazione.
- L'utilizzo di AWS CodeBuild e AWS Code Pipeline per fornire funzionalità CI/CD semplifica l'integrazione dei test di sicurezza nelle tue pipeline.
- Segui le indicazioni sulla separazione degli ambienti nel white paper Organization Your AWS Environment Using Multiple Accounts.
- Verifica l'assenza di accesso umano frequente agli ambienti in cui sono in esecuzione carichi di lavoro di produzione.
- Progetta le applicazioni in modo che supportino l'esternalizzazione dei dati di configurazione.
- Valuta se eseguire l'implementazione usando un modello di implementazione blu/verde.
- Implementa valori canary per convalidare la corretta implementazione del software.
- Utilizza strumenti crittografici come AWS Signer o AWS Key Management Service (AWS KMS) per la firma e la verifica dei pacchetti software in fase di implementazione.

Risorse

Best practice correlate:

- SEC11-BP02 Automatizza i test durante tutto il ciclo di vita di sviluppo e rilascio

Documenti correlati:

- AWS CI/CD Workshop
- Accelera le implementazioni con una governance efficace AWS
- Automatizzazione di distribuzioni pratiche e sicure
- Firma del codice tramite AWS Certificate Manager, CA privata e chiavi AWS Key Management Service asimmetriche
- Firma del codice, un controllo di fiducia e integrità per AWS Lambda

Video correlati:

- Hands-off: Automating continuous delivery pipelines at Amazon

Esempi correlati:

- Implementazioni blu/verdi con AWS Fargate

SEC11-BP07 Valutare regolarmente le proprietà di sicurezza delle condotte

Applica i principi del pilastro della sicurezza Well-Architected alle pipeline, con particolare attenzione alla separazione delle autorizzazioni. Valuta regolarmente le proprietà di sicurezza della tua infrastruttura di pipeline. Una gestione efficace della sicurezza delle pipeline assicura la protezione del software che passa attraverso le pipeline.

Risultato desiderato: le pipeline utilizzate per la creazione e implementazione del tuo software seguono le stesse pratiche consigliate di qualsiasi altro carico di lavoro nel tuo ambiente. Gli sviluppatori che li usano non possono modificare i test implementati nelle pipeline. Le pipeline devono avere solo le autorizzazioni necessarie per le implementazioni eseguite e devono applicare misure di protezione per evitare l'implementazione negli ambienti errati. Le pipeline non devono basarsi su credenziali a lungo termine e devono essere configurate in modo da emettere lo stato, per permettere la convalida dell'integrità degli ambienti di sviluppo.

Anti-pattern comuni:

- Test di sicurezza ignorabili dagli sviluppatori.
- Autorizzazioni eccessivamente elevate per le pipeline di implementazione.
- Pipeline non configurate per la convalida degli input.

- Nessuna revisione periodica delle autorizzazioni associate all'infrastruttura CI/CD.
- Uso di credenziali a lungo termine o hardcoded.

Vantaggi dell'adozione di questa best practice:

- Maggiore garanzia di integrità del software sviluppato e implementato attraverso le pipeline.
- Possibilità di arrestare un'implementazione in caso di attività sospetta.

Livello di rischio associato se questa best practice non fosse adottata: elevato

Guida all'implementazione

A partire dai servizi CI/CD gestiti che supportano i IAM ruoli si riduce il rischio di perdita delle credenziali. L'applicazione dei principi del pilastro della sicurezza all'infrastruttura di pipeline CI/CD può aiutarti a determinare i punti in cui apportare miglioramenti per la sicurezza. Seguire l'architettura di riferimento per le pipeline di implementazione AWS è un buon punto di partenza per la creazione di ambienti CI/CD. La revisione periodica dell'implementazione delle pipeline e l'analisi dei log per identificare comportamenti imprevisti può semplificare la comprensione dei modelli di utilizzo delle pipeline usate per implementare il software.

Passaggi dell'implementazione

- Parti dall'architettura di riferimento per le pipeline di implementazione AWS.
- Prendi in considerazione l'utilizzo di AWS IAMAccess Analyzer per generare in modo programmatico politiche con privilegi minimi per le pipeline. IAM
- Integra le tue pipeline con il monitoraggio e gli avvisi in modo da ricevere notifiche in caso di attività impreviste o anomale. Per AWS i servizi gestiti Amazon ti EventBridge consente di indirizzare i dati verso destinazioni come AWS LambdaAmazon Simple Notification Service (Amazon). SNS

Risorse

Documenti correlati:

- Architettura di riferimento per pipeline di implementazione AWS
- Monitoraggio di AWS CodePipeline
- Le migliori pratiche di sicurezza per AWS CodePipeline

Esempi correlati:

- DevOpsdashboard di monitoraggio (GitHub)

SEC11-BP08 Crea un programma che incorpori la titolarità della sicurezza nei team addetti ai carichi di lavoro

Crea un programma o un meccanismo che permetta ai team di sviluppo di prendere decisioni sulla sicurezza del software che creano. Il team della sicurezza dovrà convalidare queste decisioni durante una revisione, ma integrare la proprietà della sicurezza nei team di sviluppo consente di creare carichi di lavoro più veloci e sicuri. Questo meccanismo promuove anche una cultura della responsabilità che ha un impatto positivo sul funzionamento dei sistemi che crei.

Risultato desiderato: per integrare titolarità e processo decisionale in materia di sicurezza nei team di sviluppo, è possibile formare gli sviluppatori in merito a come pensare alla sicurezza o migliorare la loro formazione integrando personale addetto alla sicurezza o associato ai team di sviluppo. Entrambi gli approcci sono validi e permettono al team di prendere decisioni di qualità migliore sulla sicurezza nelle fasi iniziali del ciclo di sviluppo. Questo modello di proprietà è basato sulla formazione per la sicurezza delle applicazioni. Iniziando dal modello di rischio per il carico di lavoro specifico, puoi concentrarti sul design thinking nel contesto appropriato. Un altro vantaggio della presenza di una comunità di sviluppatori attenti alla sicurezza, o di un gruppo di tecnici della sicurezza che collabora con i team di sviluppo, è la possibilità di comprendere a pieno il modo in cui è compilato il codice. Questa comprensione permette di determinare le aree di miglioramento successive per l'automazione.

Anti-pattern comuni:

- Assegnazione di tutte le decisioni in materia di sicurezza al team responsabile della sicurezza.
- Gestione dei requisiti di sicurezza in fasi tardive del processo di sviluppo.
- Assenza di feedback di sviluppatori e responsabili della sicurezza sul funzionamento del programma.

Vantaggi dell'adozione di questa best practice:

- Riduzione del tempo necessario per completare le revisioni della sicurezza.
- Riduzione dei problemi di sicurezza rilevati solo in fase di revisione della sicurezza.
- Miglioramento della qualità complessiva del software compilato.

- Opportunità di identificare e comprendere i problemi sistematici o le aree di miglioramento a valore elevato.
- Riduzione della quantità di attività di correzione dovute agli esiti delle revisioni della sicurezza.
- Migliore percezione della funzione della sicurezza.

Livello di rischio associato se questa best practice non fosse adottata: basso

Guida all'implementazione

Per iniziare, attieniti alle linee guida illustrate in SEC11-BP01 Train per la sicurezza delle applicazioni Identifica quindi il modello operativo per il programma che ritieni più efficace per l'organizzazione. I due modelli principali consistono nel formare gli sviluppatori o nell'integrare responsabili della sicurezza nei team di sviluppo. Una volta scelto l'approccio iniziale, devi eseguire un progetto pilota con un singolo team o un piccolo gruppo di team del carico di lavoro per dimostrare il funzionamento del modello per l'organizzazione. Il supporto autorevole da parte dello sviluppatore e di altre parti responsabili della sicurezza dell'organizzazione semplifica l'implementazione e il successo del programma. Durante la creazione del programma, è importante scegliere le metriche da usare per dimostrarne il valore. Imparare da come AWS ha affrontato questo problema è una buona esperienza di apprendimento. Questa best practice è per lo più incentrata sulla trasformazione e sulla cultura aziendali. Gli strumenti usati devono supportare la collaborazione tra lo sviluppatore e le comunità responsabili della sicurezza.

Passaggi dell'implementazione

- Per iniziare, predisponi corsi di formazione sulla sicurezza delle applicazioni per gli sviluppatori.
- Crea una community e un programma di onboarding per formare gli sviluppatori.
- Scegli un nome per il programma. Alcuni termini comunemente usati sono Responsabilità, Supporto o Promozione.
- Identifica il modello da usare: formazione per gli sviluppatori, integrazione di tecnici della sicurezza o ruoli di sicurezza per affinità.
- Identifica alcuni sponsor del progetto tra responsabili della sicurezza, sviluppatori e altri gruppi potenzialmente pertinenti.
- Tieni traccia delle metriche per il numero di persone coinvolte nel programma, del tempo impiegato per le revisioni e del feedback ottenuto da sviluppatori e responsabili della sicurezza. Usa queste metriche per apportare miglioramenti.

Risorse

Best practice correlate:

- SEC11-BP01 Train per la sicurezza delle applicazioni
- SEC11-BP02 Automatizza i test durante tutto il ciclo di vita di sviluppo e rilascio

Documenti correlati:

- How to approach threat modeling
- How to think about cloud security governance

Video correlati:

- Proactive security: Considerations and approaches

Affidabilità

Il pilastro dell'affidabilità comprende la capacità di un carico di lavoro di eseguire la funzione attesa in modo corretto e coerente quando previsto. Puoi trovare linee guida prescrittive sull'implementazione nel whitepaper sul pilastro dell'affidabilità.

Aree delle best practice

- Fondamenti
- Architettura del carico di lavoro
- Gestione delle modifiche
- Gestione dei guasti

Fondamenti

Questions

- REL1. Come si gestiscono Service Quotas e restrizioni?
- REL2. Come si pianifica la topologia della rete?

REL1. Come si gestiscono Service Quotas e restrizioni?

Per le architetture di carichi di lavoro basate sul cloud, esistono Service Quotas (chiamate anche restrizioni dei servizi). Queste quote servono a prevenire l'approvvigionamento accidentale di più risorse del necessario e a limitare la frequenza delle richieste relative alle API operazioni in modo da proteggere i servizi da eventuali abusi. Esistono anche vincoli di risorse, ad esempio la velocità con cui è possibile trasferire i bit su un cavo in fibra ottica o lo spazio di archiviazione su un disco fisico.

Best practice

- REL01-BP01 Conoscenza delle quote e dei vincoli di servizio
- REL01-BP02 Gestisci le quote di servizio tra account e regioni
- REL01-BP03 Soddisfa quote e vincoli di servizio fissi tramite l'architettura
- REL01-BP04 Monitoraggio e gestione delle quote
- REL01-BP05 Automatizza la gestione delle quote
- REL01-BP06 Assicurarsi che esista un intervallo sufficiente tra le quote correnti e l'utilizzo massimo per consentire il failover

REL01-BP01 Conoscenza delle quote e dei vincoli di servizio

Conosci le quote predefinite e gestisci le richieste di aumento delle quote per l'architettura del carico di lavoro. Sai quali vincoli delle risorse cloud, ad esempio disco o rete, sono potenzialmente influenti.

Risultato desiderato: i clienti possono prevenire il degrado o l'interruzione del servizio Account AWS implementando linee guida appropriate per il monitoraggio delle metriche chiave, le revisioni dell'infrastruttura e le misure correttive dell'automazione per verificare che non vengano raggiunte le quote e i vincoli di servizio che potrebbero causare il degrado o l'interruzione del servizio.

Anti-pattern comuni:

- Distribuzione di un carico di lavoro senza comprendere le quote hard o soft e i relativi limiti per i servizi utilizzati.
- Distribuzione di un carico di lavoro sostitutivo senza analizzare e riconfigurare le quote necessarie o contattare preventivamente l'assistenza.
- Supposizione che i servizi cloud non abbiano limiti e che i servizi possano essere utilizzati senza tener conto di tariffe, limiti, conteggi, quantità.
- Supposizione che le quote verranno aumentate automaticamente.

- Mancata conoscenza del processo e della scadenza delle richieste di quote.
- Supposizione che la quota predefinita del servizio cloud sia identica per ogni servizio rispetto alle varie regioni.
- Supposizione che i vincoli del servizio possano essere violati e che i sistemi procedano al dimensionamento automatico o aumentino il limite oltre i vincoli della risorsa
- Nessun test dell'applicazione nei momenti di picco del traffico, per stressare l'utilizzo delle sue risorse.
- Provisioning della risorsa senza analisi della dimensione della risorsa richiesta.
- Provisioning in eccesso di capacità scegliendo tipi di risorse che vanno ben oltre il fabbisogno effettivo o i picchi previsti.
- Nessuna valutazione dei requisiti di capacità per nuovi livelli di traffico prima di un nuovo evento cliente o dell'implementazione di una nuova tecnologia.

Vantaggi dell'adozione di questa best practice: il monitoraggio e la gestione automatizzata di quote di servizio e vincoli di risorse consentono di ridurre in modo proattivo i guasti. Le modifiche nei modelli di traffico per il servizio di un cliente possono causare un'interruzione o un degrado se non si seguono le best practice. Monitorando e gestendo questi valori in tutte le regioni e in tutti gli account, le applicazioni possono avere una maggiore resilienza in caso di eventi avversi o non pianificati.

Livello di rischio associato se questa best practice non fosse adottata: elevato

Guida all'implementazione

Service Quotas è un AWS servizio che ti aiuta a gestire le quote per oltre 250 AWS servizi da un'unica posizione. Oltre a cercare i valori delle quote, puoi anche richiedere e tenere traccia degli aumenti delle quote dalla console Service Quotas o utilizzando il. AWS SDK AWS Trusted Advisor offre un controllo delle quote di servizio che mostra l'utilizzo e le quote per alcuni aspetti di alcuni servizi. Le quote di servizio predefinite per servizio si trovano anche nella AWS documentazione del rispettivo servizio (ad esempio, vedi Amazon VPC Quotas).

Alcuni limiti di servizio, come i limiti di velocità per le limitazioni di velocità, APIs vengono impostati all'interno dello stesso Amazon API Gateway configurando un piano di utilizzo. Alcuni limiti impostati come configurazione sui rispettivi servizi includono ProvisionedIOPS, Amazon RDS storage allocated e Amazon EBS Volume Allocations. Amazon Elastic Compute Cloud dispone di un proprio pannello di controllo sulle restrizioni dei servizi che consente di gestire l'istanza, Amazon Elastic Block Store e i limiti degli indirizzi IP elastici. Se hai un caso d'uso in cui le quote di servizio influiscono sulle

prestazioni dell'applicazione e non sono adattabili alle tue esigenze, contattaci per verificare se esistono soluzioni AWS Support di mitigazione.

Le quote di servizio possono essere specifiche per ogni regione o di natura globale. L'utilizzo AWS di un servizio che raggiunge la quota prevista non funzionerà come previsto in condizioni di utilizzo normale e potrebbe causare interruzioni o deterioramenti del servizio. Ad esempio, una quota di servizio limita il numero di EC2 istanze Amazon DL utilizzate in una regione. Tale limite può essere raggiunto durante un evento di scalabilità del traffico utilizzando i gruppi ASG Auto Scaling ().

Le quote di servizio per ogni account devono essere valutate in modo regolare per determinare i limiti di servizio opportuni per quell'account. Queste quote di servizio fungono da guardrail operativi, per evitare di fornire accidentalmente più risorse di quelle necessarie. Servono anche a limitare i tassi di richiesta sulle API operazioni per proteggere i servizi dagli abusi.

I limiti dei servizi sono diversi dalle quote dei servizi. I vincoli di servizio rappresentano i limiti di una particolare risorsa, definiti dalla stessa. Questi potrebbero essere la capacità di archiviazione (ad esempio, gp2 ha un limite di dimensione compreso tra 1 GB e 16 TB) o la velocità effettiva del disco. È essenziale che il vincolo di un tipo di risorsa sia progettato e valutato in modo costante per l'utilizzo che potrebbe raggiungere il suo limite. In caso di raggiungimento inaspettato di un vincolo, può verificarsi il degrado o l'interruzione delle applicazioni o dei servizi dell'account.

Se c'è un caso d'uso in cui le quote di servizio influiscono sulle prestazioni di un'applicazione e non possono essere adattate alle esigenze richieste, contattateci AWS Support per verificare se esistono delle mitigazioni. Per ulteriori dettagli sull'adeguamento delle quote fisse, consulta REL01-BP03 Soddisfa quote e vincoli di servizio fissi tramite l'architettura.

Esistono numerosi AWS servizi e strumenti per aiutare a monitorare e gestire Service Quotas. Sfrutta il servizio e gli strumenti per fornire controlli automatizzati o manuali dei livelli di quota.

- AWS Trusted Advisor offre un controllo delle quote di servizio che mostra l'utilizzo e le quote per alcuni aspetti di alcuni servizi. Può aiutare a identificare i servizi vicini alle quote.
- AWS Management Console fornisce metodi per visualizzare i valori delle quote dei servizi, gestire, richiedere nuove quote, monitorare lo stato delle richieste di quote e visualizzare la cronologia delle quote.
- AWS CLI e CDKs offre metodi programmatici per gestire e monitorare automaticamente i livelli e l'utilizzo delle quote di servizio.

Passaggi dell'implementazione

Per Service Quotas:

- Rivedi AWS Service Quotas.
- Per conoscere le quote di servizio esistenti, determina i servizi (come IAM Access Analyzer) da utilizzare. Esistono circa 250 AWS servizi controllati da quote di servizio. Quindi stabilisci il nome della quota di servizio specifica utilizzabile all'interno di ogni account e regione. Esistono circa 3000 nomi di quote di servizio per regione.
- Amplia questa analisi delle quote AWS Config per trovare tutte le AWS risorse utilizzate nel tuo. Account AWS
- Usa AWS CloudFormation i dati per determinare le AWS risorse utilizzate. Guarda le risorse che sono state create con AWS Management Console o con il `list-stack-resources` AWS CLI comando. È anche possibile vedere le risorse configurate da implementare nel modello stesso.
- Stabilisci tutti i servizi necessari per il tuo carico di lavoro analizzando il codice di implementazione.
- Determina le quote di servizio applicabili. Utilizza le informazioni accessibili a livello di programmazione da Trusted Advisor e Service Quotas.
- Stabilisci un metodo di monitoraggio automatizzato (consulta REL01-BP02 Gestisci le quote di servizio tra account e regioni e REL01-BP04 Monitoraggio e gestione delle quote) per ricevere avvisi e informazioni se le quote di servizio sono vicine al limite o lo hanno superato.
- Stabilisci un metodo automatizzato e programmatico per verificare se una quota di servizio ha subito modifiche in una regione ma non in altre nello stesso account (consulta REL01-BP02 Gestisci le quote di servizio tra account e regioni e REL01-BP04 Monitoraggio e gestione delle quote).
- Automatizza la scansione dei log e delle metriche delle applicazioni per determinare la presenza di errori di quota o di vincoli di servizio. In presenza di errori, invia gli allarmi al sistema di monitoraggio.
- Stabilisci procedure di progettazione per calcolare la modifica richiesta della quota (consulta REL01-BP05 Automatizza la gestione delle quote) una volta individuata la necessità di quote più elevate per servizi specifici.
- Crea un flusso di lavoro di provisioning e di approvazione per richiedere modifiche alla quota di servizio, che dovrebbe includere un flusso di lavoro di eccezione in caso di rifiuto della richiesta o di approvazione parziale.
- Crea un metodo ingegneristico per rivedere le quote di servizio prima di fornire e utilizzare nuovi AWS servizi prima di implementarli in ambienti di produzione o caricati. (ad esempio, account per il test di carico).

Per i vincoli dei servizi:

- Stabilisci metodi di monitoraggio e metriche per avvisi in caso di avvicinamento da parte delle risorse ai relativi limiti. Sfruttalo in CloudWatch modo appropriato per le metriche o il monitoraggio dei log.
- Stabilisci soglie di allarme per ciascuna risorsa con un vincolo significativo per l'applicazione o il sistema.
- Crea procedure di gestione del flusso di lavoro e dell'infrastruttura per cambiare il tipo di risorsa se il vincolo è prossimo all'utilizzo. Questo flusso di lavoro dovrebbe includere test di carico come best practice per verificare che quello nuovo sia il tipo di risorsa corretto in base ai nuovi vincoli.
- Migra la risorsa identificata al nuovo tipo di risorsa consigliato, utilizzando procedure e processi esistenti.

Risorse

Best practice correlate:

- REL01-BP02 Gestisci le quote di servizio tra account e regioni
- REL01-BP03 Soddisfa quote e vincoli di servizio fissi tramite l'architettura
- REL01-BP04 Monitoraggio e gestione delle quote
- REL01-BP05 Automatizza la gestione delle quote
- REL01-BP06 Assicurarsi che esista un intervallo sufficiente tra le quote correnti e l'utilizzo massimo per consentire il failover
- REL03-BP01 Scegli come segmentare il tuo carico di lavoro
- REL10-BP01 Implementazione del carico di lavoro in più sedi
- REL11-BP01 Monitora tutti i componenti del carico di lavoro per rilevare i guasti
- REL11-BP03 Automatizza la guarigione su tutti i livelli
- REL12-BP05 Verifica la resilienza utilizzando l'ingegneria del caos

Documenti correlati:

- AWS Il pilastro dell'affidabilità di Well-Architected Framework: disponibilità
- AWS Service Quotas (precedentemente denominate limiti di servizio)
- AWS Trusted Advisor Controlli relativi alle migliori pratiche (vedere la sezione Limiti del servizio)

- AWS limita il monitoraggio delle AWS risposte
- Limiti EC2 del servizio Amazon
- What is Service Quotas?
- How to Request Quota Increase
- Service endpoints and quotas
- Guida per l'utente di Service Quotas
- Quota Monitor per AWS
- AWS Limiti di isolamento dei guasti
- Availability with redundancy
- AWS per i dati
- Cos'è l'integrazione continua?
- Cos'è la distribuzione continua?
- APNPartner: partner che possono aiutare nella gestione della configurazione
- Gestione del ciclo di vita dell'account in ambienti SaaS account-per-tenant su AWS
- Gestione e monitoraggio della limitazione dei carichi di API lavoro
- Visualizza i AWS Trusted Advisor consigli su larga scala con AWS Organizations
- Automatizzazione degli aumenti dei limiti di servizio e del supporto aziendale con AWS Control Tower

Video correlati:

- AWS Live re:InForce 2019 - Service Quotas
- Visualizzazione e gestione delle quote per AWS i servizi tramite Service Quotas
- AWS IAMDimostrazione di Quotas

Strumenti correlati:

- CodeGuru Revisore Amazon
- AWS CodeDeploy
- AWS CloudTrail
- Amazon CloudWatch

- Amazon EventBridge
- Amazon DevOps Guru
- AWS Config
- AWS Trusted Advisor
- AWS CDK
- AWS Systems Manager
- Marketplace AWS

REL01-BP02 Gestisci le quote di servizio tra account e regioni

Se utilizzi più account o regioni, assicurati di richiedere le quote opportune in tutti gli ambienti di esecuzione dei carichi di lavoro di produzione.

Risultato desiderato: servizi e applicazioni non dovrebbero essere influenzati dall'esaurimento della quota di servizio per le configurazioni che si estendono su account o regioni o che presentano progetti di resilienza che utilizzano il failover di zona, regione o account.

Anti-pattern comuni:

- Si consente l'aumento dell'utilizzo delle risorse in una regione di isolamento senza alcun meccanismo per mantenere la capacità nelle altre.
- Impostazione manualmente tutte le quote in modo indipendente nelle regioni di isolamento.
- Mancata valutazione dell'effetto delle architetture di resilienza (come quelle attive o passive) nelle future esigenze di quote durante un degrado nella regione non primaria.
- Mancata valutazione regolare delle quote e applicazione dele modifiche necessarie in ogni regione e account in cui viene gestito il carico di lavoro.
- Mancato utilizzo dei modelli di richiesta di quote per la richiesta di aumenti in più regioni e account.
- Mancato aggiornamento delle quote dei servizi, perché si pensa erroneamente che l'aumento delle quote abbia implicazioni di costo, come le richieste di prenotazione di calcolo.

Vantaggi dell'adozione di questa best practice: verifica della capacità di gestire il carico corrente nelle regioni o negli account secondari in caso di indisponibilità dei servizi regionali. Questo consente di ridurre il numero di errori o livelli di degrado che si verificano durante la perdita di regioni.

Livello di rischio associato se questa best practice non fosse adottata: elevato

Guida all'implementazione

Il monitoraggio delle quote di servizio avviene per account. Salvo diversa indicazione, ogni quota è specifica. Regione AWS Oltre agli ambienti di produzione, gestisci anche le quote in tutti gli ambienti non di produzione applicabili, in modo che test e sviluppo non siano ostacolati. Il mantenimento di un elevato grado di resilienza richiede una valutazione continua delle quote di servizio (sia automatica che manuale).

Con un aumento dei carichi di lavoro in tutte le regioni dovuto all'implementazione di progetti che utilizzano approcci attivo/attivo, attivo/passivo con standby a caldo, attivo/passivo con standby a freddo e attivo/passivo con Pilot Light, è essenziale conoscere tutti i livelli di quota di regione e account. I modelli di traffico passati non sono sempre un buon indicatore per stabilire se la quota di servizio è impostata correttamente.

Altrettanto importante è il fatto che il limite di nome della quota di servizio non è sempre lo stesso per ogni regione. In una regione, il valore potrebbe essere cinque, in un'altra potrebbe essere dieci. La gestione di queste quote deve riguardare tutti gli stessi servizi, account e regioni per garantire una resilienza costante sotto carico.

Riconcilia tutte le differenze di quota di servizio tra le diverse regioni (regione attiva o passiva) e crea processi per riconciliare continuamente queste differenze. I piani di test dei failover passivi delle regioni sono raramente scalati in base alla capacità attiva di picco, il che significa che gli esercizi delle giornate di gioco o table top potrebbero non riuscire a trovare le differenze nelle quote di servizio tra le regioni e a mantenere i limiti corretti.

La deviazione della quota di servizio, la condizione in cui la modifica dei limiti della quota di servizio per una determinata quota denominata avviene in una regione e non in tutte le regioni, è un fattore molto importante da monitorare e valutare. Si dovrebbe prendere in considerazione la possibilità di modificare la quota nelle regioni con traffico o potenzialmente in grado di trasportare traffico.

- Seleziona account e regioni pertinenti in base ai tuoi requisiti di servizio, latenza, normativi e ripristino di emergenza.
- Identifica le quote dei servizi per tutti gli account, le regioni e le zone di disponibilità pertinenti. Le restrizioni si riferiscono ad account e regione. Confronta questi valori per individuare le differenze.

Passaggi dell'implementazione

- Rivedi i valori di Service Quotas che potrebbero aver superato il livello di rischio di utilizzo. AWS Trusted Advisor offre allarmi per la violazione di soglie dell'80% e del 90%.

- Rivedi i valori per le quote di servizio in qualsiasi regione passiva (in un progetto Attivo/Passivo). Verifica che il carico venga eseguito in modo corretto nelle regioni secondarie in caso di guasto nella regione primaria.
- Valuta in modo automatizzato se si è verificata una deviazione delle quote di servizio tra le regioni dello stesso account e agisci di conseguenza per modificare i limiti.
- Se le unità organizzative (UO) del cliente sono strutturate nel modo supportato, aggiorna i modelli di quote di servizio per riflettere le modifiche alle quote da applicare a più regioni e account.
  - Crea un modello e associa le regioni alla modifica della quota.
  - Rivedi tutti i modelli delle quote di servizio esistenti per qualsiasi modifica richiesta (regione, limiti e account).

Risorse

Best practice correlate:

- REL01-BP01 Conoscenza delle quote e dei vincoli di servizio
- REL01-BP03 Soddisfa quote e vincoli di servizio fissi tramite l'architettura
- REL01-BP04 Monitoraggio e gestione delle quote
- REL01-BP05 Automatizza la gestione delle quote
- REL01-BP06 Assicurarsi che esista un intervallo sufficiente tra le quote correnti e l'utilizzo massimo per consentire il failover
- REL03-BP01 Scegli come segmentare il tuo carico di lavoro
- REL10-BP01 Implementazione del carico di lavoro in più sedi
- REL11-BP01 Monitora tutti i componenti del carico di lavoro per rilevare i guasti
- REL11-BP03 Automatizza la guarigione su tutti i livelli
- REL12-BP05 Verifica la resilienza utilizzando l'ingegneria del caos

Documenti correlati:

- AWS Il pilastro dell'affidabilità di Well-Architected Framework: disponibilità
- AWS Service Quotas (precedentemente denominate limiti di servizio)
- AWS Trusted Advisor Controlli relativi alle migliori pratiche (vedere la sezione Limiti del servizio)
- AWS limita il monitoraggio delle AWS risposte

- Limiti EC2 del servizio Amazon
- What is Service Quotas?
- How to Request Quota Increase
- Service endpoints and quotas
- Guida per l'utente di Service Quotas
- Quota Monitor per AWS
- AWS Limiti di isolamento dei guasti
- Availability with redundancy
- AWS per i dati
- Cos'è l'integrazione continua?
- Cos'è la distribuzione continua?
- APNPartner: partner che possono aiutare nella gestione della configurazione
- Gestione del ciclo di vita dell'account in ambienti SaaS account-per-tenant su AWS
- Gestione e monitoraggio della limitazione dei carichi di API lavoro
- Visualizza i AWS Trusted Advisor consigli su larga scala con AWS Organizations
- Automatizzazione degli aumenti dei limiti di servizio e del supporto aziendale con AWS Control Tower

Video correlati:

- AWS Live re:INforce 2019 - Service Quotas
- Visualizzazione e gestione delle quote per AWS i servizi tramite Service Quotas
- AWS IAMDemo di Quotas

Servizi correlati:

- CodeGuru Revisore Amazon
- AWS CodeDeploy
- AWS CloudTrail
- Amazon CloudWatch
- Amazon EventBridge

- Amazon DevOps Guru
- AWS Config
- AWS Trusted Advisor
- AWS CDK
- AWS Systems Manager
- Marketplace AWS

REL01-BP03 Soddisfa quote e vincoli di servizio fissi tramite l'architettura

Identifica con attenzione quote di servizio, vincoli del servizio e limiti delle risorse fisiche che non possono essere modificati. Progetta architetture per applicazioni e servizi in modo da impedire che questi limiti pregiudichino l'affidabilità.

Gli esempi includono la larghezza di banda di rete, la dimensione del payload per l'invocazione di funzioni senza server, la velocità di accelerazione per un gateway e le connessioni simultanee degli utenti a un database. API

Risultato desiderato: funzionamento dell'applicazione o del servizio come previsto in condizioni di traffico normale e intenso. L'applicazione o il servizio è stato progettato per operare entro i limiti dei vincoli o delle quote di servizio fissi della risorsa.

Anti-pattern comuni:

- Scelta di una progettazione che usa una risorsa di un servizio, senza essere al corrente della presenza di vincoli che causeranno errori di progettazione durante il dimensionamento.
- Esecuzione di benchmark poco realistici e che raggiungono le quote di servizio fisse durante i test. Ad esempio, l'esecuzione di test a un limite di espansione per un periodo di tempo prolungato.
- Scelta di una progettazione non scalabile o modificabile in caso di superamento delle quote di servizio fisse. Ad esempio, una dimensione del payload di 256 KB. SQS
- Mancata progettazione e implementazione dell'osservabilità per monitorare e inviare avvisi circa le soglie per le quote di servizio a rischio durante eventi di traffico elevato.

Vantaggi dell'adozione di questa best practice: verifica del funzionamento dell'applicazione con tutti i livelli di carico dei servizi previsti senza interruzioni o deterioramenti.

Livello di rischio associato se questa best practice non fosse adottata: medio

Guida all'implementazione

A differenza delle quote o delle risorse di soft service che vengono sostituite con unità di capacità superiore, le quote fisse AWS dei servizi non possono essere modificate. Ciò significa che tutti questi tipi di AWS servizi devono essere valutati per i potenziali limiti di capacità rigida quando vengono utilizzati nella progettazione di un'applicazione.

I limiti fissi vengono mostrati nella console di Service Quotas. Se le colonne visualizzano `ADJUSTABLE = No`, il servizio ha un limite fisso. I limiti fissi vengono mostrati anche in alcune pagine di configurazione delle risorse. Ad esempio, Lambda presenta un limite fisso specifico che non può essere modificato.

Ad esempio, durante la progettazione di un'applicazione Python da eseguire in una funzione Lambda, l'applicazione deve essere valutata per determinare la probabilità di un'esecuzione di Lambda superiore a 15 minuti. Se il codice potrebbe restare in esecuzione oltre questo limite della quota di servizio, devi prendere in considerazione tecnologie o progettazioni alternative. In caso di raggiungimento del limite dopo l'implementazione nell'ambiente di produzione, l'applicazione sarà soggetta a errori o interruzioni fino alla correzione. A differenza dalle quote flessibili, non esiste alcun metodo per modificare i limiti, anche in caso di eventi di emergenza con livello di gravità 1.

Una volta implementata l'applicazione in un ambiente di test, occorre adottare una strategia per determinare se l'eventuale probabilità di raggiungere i limiti fissi. I test di stress, di carico e di caos devono fare parte del piano di test iniziale.

Passaggi dell'implementazione

- Consulta l'elenco completo dei AWS servizi che potrebbero essere utilizzati nella fase di progettazione dell'applicazione.
- Esamina i limiti di quota flessibili e fissi per tutti i servizi. Non tutti i limiti vengono mostrati nella console di Service Quotas. Alcuni servizi indicano tali limiti in posizioni alternative.
- Nel progettare l'applicazione, esamina i principali fattori commerciali e tecnologici del carico di lavoro, come risultati aziendali, casi d'uso, sistemi dipendenti, obiettivi di disponibilità e oggetti di ripristino di emergenza. Orienta il processo di identificazione del sistema distribuito corretto per il carico di lavoro in base a tali fattori commerciali e tecnologici.
- Analizza il carico dei servizi tra regioni e account. Molti limiti fissi per i servizi variano a seconda della regione. Tuttavia, alcuni limiti dipendono dagli account.
- Analizza le architetture di resilienza per l'utilizzo delle risorse durante un guasto a livello di zona e di regione. Nel corso dello sviluppo di progettazioni multi-regione che usano approcci attivo/

attivo, attivo/passivo con standby a caldo, attivo/passivo con standby a freddo e attivo/passivo con Pilot Light, i casi di errore determineranno un utilizzo più elevato. Questo comportamento crea un possibile caso d'uso per il raggiungimento dei limiti fissi.

Risorse

Best practice correlate:

- REL01-BP01 Conoscenza delle quote e dei vincoli di servizio
- REL01-BP02 Gestisci le quote di servizio tra account e regioni
- REL01-BP04 Monitoraggio e gestione delle quote
- REL01-BP05 Automatizza la gestione delle quote
- REL01-BP06 Assicurarsi che esista un intervallo sufficiente tra le quote correnti e l'utilizzo massimo per consentire il failover
- REL03-BP01 Scegli come segmentare il tuo carico di lavoro
- REL10-BP01 Implementazione del carico di lavoro in più sedi
- REL11-BP01 Monitora tutti i componenti del carico di lavoro per rilevare i guasti
- REL11-BP03 Automatizza la guarigione su tutti i livelli
- REL12-BP05 Verifica la resilienza utilizzando l'ingegneria del caos

Documenti correlati:

- AWS Il pilastro dell'affidabilità di Well-Architected Framework: disponibilità
- AWS Service Quotas (precedentemente denominate limiti di servizio)
- AWS Trusted Advisor Controlli relativi alle migliori pratiche (vedere la sezione Limiti del servizio)
- AWS limita il monitoraggio delle AWS risposte
- Limiti EC2 del servizio Amazon
- What is Service Quotas?
- How to Request Quota Increase
- Service endpoints and quotas
- Guida per l'utente di Service Quotas
- Quota Monitor per AWS

- AWS Limiti di isolamento dei guasti
- Availability with redundancy
- AWS per i dati
- Cos'è l'integrazione continua?
- Cos'è la distribuzione continua?
- APNPartner: partner che possono aiutare nella gestione della configurazione
- Gestione del ciclo di vita dell'account in ambienti SaaS account-per-tenant su AWS
- Gestione e monitoraggio della limitazione dei carichi di API lavoro
- Visualizza i AWS Trusted Advisor consigli su larga scala con AWS Organizations
- Automatizzazione degli aumenti dei limiti di servizio e del supporto aziendale con AWS Control Tower
- Actions, resources, and condition keys for Service Quotas

Video correlati:

- AWS Live re:InForce 2019 - Service Quotas
- Visualizzazione e gestione delle quote per AWS i servizi tramite Service Quotas
- AWS IAMDimostrazione di Quotas
- AWS re:Invent 2018: Chiudere i circuiti e aprire le menti: come assumere il controllo di sistemi, grandi e piccoli

Strumenti correlati:

- AWS CodeDeploy
- AWS CloudTrail
- Amazon CloudWatch
- Amazon EventBridge
- Amazon DevOps Guru
- AWS Config
- AWS Trusted Advisor
- AWS CDK
- AWS Systems Manager

- Marketplace AWS

REL01-BP04 Monitoraggio e gestione delle quote

Valuta il tuo utilizzo potenziale e aumenta le quote in modo opportuno per una crescita pianificata dell'utilizzo.

Risultato desiderato: implementazione di sistemi attivi e automatizzati per la gestione e il monitoraggio. Queste soluzioni operative indicano che le soglie di utilizzo delle quote stanno per essere raggiunte. Il problema può essere risolto in modo proattivo tramite modifiche alle quote richieste.

Anti-pattern comuni:

- Mancata configurazione del monitoraggio per verificare le soglie delle quote di servizio.
- Mancata configurazione del monitoraggio dei limiti fissi, anche se i valori non possono essere modificati.
- Valutazione errata della quantità di tempo necessaria per richiedere e ottenere la modifica di una quota flessibile, supponendo che sia immediata o rapida.
- Configurazione di allarmi per l'avvicinamento alle quote di servizio, ma senza alcun processo di risposta a un avviso.
- Configurazione degli allarmi solo per i servizi supportati da Service AWS Quotas e non monitoraggio di altri servizi. AWS
- Valutazione errata della gestione delle quote per progettazioni di resilienza in più regioni, come gli approcci attivo/attivo, attivo/passivo con standby a caldo, attivo/passivo con standby a freddo e attivo/passivo con Pilot Light.
- Mancata valutazione delle differenze di quota tra regioni.
- Mancata valutazione delle esigenze in ogni regione per una richiesta di aumento di quota specifica.
- Mancato utilizzo di modelli per la gestione delle quote multiregioni.

Vantaggi derivanti dall'adozione di questa best practice: il monitoraggio automatico delle AWS Service Quotas e il monitoraggio dell'utilizzo rispetto a tali quote ti consentiranno di vedere quando ti stai avvicinando a un limite di quota. Puoi usare questi dati di monitoraggio per limitare eventuali errori dovuti all'esaurimento della quota.

Livello di rischio associato se questa best practice non fosse adottata: medio

Guida all'implementazione

Per i servizi supportati, puoi monitorare le quote configurando servizi diversi in grado di eseguire una valutazione e quindi inviare avvisi o allarmi. In questo modo, il monitoraggio dell'utilizzo è più semplice e puoi ricevere avvisi all'avvicinamento delle quote. Questi allarmi possono essere richiamati da, funzioni AWS Config Lambda, Amazon CloudWatch o da. AWS Trusted Advisor Puoi anche utilizzare i filtri metrici sui CloudWatch registri per cercare ed estrarre modelli nei log per determinare se l'utilizzo si avvicina alle soglie di quota.

Passaggi dell'implementazione

Per il monitoraggio:

- Acquisisci informazioni sull'attuale consumo di risorse, ad esempio bucket o istanze. Utilizza API le operazioni di servizio, come Amazon EC2 `DescribeInstances`API, per raccogliere l'attuale consumo di risorse.
- Acquisisci le attuali quote essenziali e valide per i servizi usando:
  - AWS Service Quotas
  - AWS Trusted Advisor
  - AWS documentazione
  - AWS pagine specifiche del servizio
  - AWS Command Line Interface (AWS CLI)
  - AWS Cloud Development Kit (AWS CDK)
- Utilizza AWS Service Quotas, un AWS servizio che ti aiuta a gestire le quote per oltre 250 AWS servizi da un'unica posizione.
- Utilizza i limiti Trusted Advisor di servizio per monitorare i tuoi attuali limiti di servizio a varie soglie.
- Utilizza la cronologia delle quote di servizio (console o AWS CLI) per verificare gli aumenti regionali.
- Confronta la modifica delle quote di servizio in ogni regione e ogni account per creare equivalenze, se necessario.

Per la gestione:

- Automatizzato: imposta una regola AWS Config personalizzata per scansionare le quote di servizio tra le regioni e confrontarle per individuare le differenze.

- Automatica: configura una funzione Lambda personalizzata per analizzare le quote di servizio tra regioni e confrontarle per individuare le differenze.
- Manuale: scansiona la quota dei servizi tramite AWS CLI o AWS Console per scansionare le quote di servizio tra le regioni e confrontarle per individuare eventuali differenze. API Segnala le differenze.
- In caso di individuazione di differenze nelle quote tra regioni, richiedi una modifica della quota, se necessario.
- Esamina il risultato di tutte le richieste.

Risorse

Best practice correlate:

- REL01-BP01 Conoscenza delle quote e dei vincoli di servizio
- REL01-BP02 Gestisci le quote di servizio tra account e regioni
- REL01-BP03 Soddisfa quote e vincoli di servizio fissi tramite l'architettura
- REL01-BP05 Automatizza la gestione delle quote
- REL01-BP06 Assicurarsi che esista un intervallo sufficiente tra le quote correnti e l'utilizzo massimo per consentire il failover
- REL03-BP01 Scegli come segmentare il tuo carico di lavoro
- REL10-BP01 Implementazione del carico di lavoro in più sedi
- REL11-BP01 Monitora tutti i componenti del carico di lavoro per rilevare i guasti
- REL11-BP03 Automatizza la guarigione su tutti i livelli
- REL12-BP05 Verifica la resilienza utilizzando l'ingegneria del caos

Documenti correlati:

- AWS Il pilastro dell'affidabilità di Well-Architected Framework: disponibilità
- AWS Service Quotas (precedentemente denominate limiti di servizio)
- AWS Trusted Advisor Controlli relativi alle migliori pratiche (vedere la sezione Limiti del servizio)
- AWS limita il monitoraggio delle AWS risposte
- Limiti EC2 del servizio Amazon
- What is Service Quotas?

- How to Request Quota Increase
- Service endpoints and quotas
- Guida per l'utente di Service Quotas
- Quota Monitor per AWS
- AWS Limiti di isolamento dei guasti
- Availability with redundancy
- AWS per i dati
- Cos'è l'integrazione continua?
- Cos'è la distribuzione continua?
- APNPartner: partner che possono aiutare nella gestione della configurazione
- Gestione del ciclo di vita dell'account in ambienti SaaS account-per-tenant su AWS
- Gestione e monitoraggio della limitazione dei carichi di API lavoro
- Visualizza i AWS Trusted Advisor consigli su larga scala con AWS Organizations
- Automatizzazione degli aumenti dei limiti di servizio e del supporto aziendale con AWS Control Tower
- Actions, resources, and condition keys for Service Quotas

Video correlati:

- AWS Live re:InForce 2019 - Service Quotas
- Visualizzazione e gestione delle quote per AWS i servizi tramite Service Quotas
- AWS IAMDimostrazione di Quotas
- AWS re:Invent 2018: Chiudere i circuiti e aprire le menti: come assumere il controllo di sistemi, grandi e piccoli

Strumenti correlati:

- AWS CodeDeploy
- AWS CloudTrail
- Amazon CloudWatch
- Amazon EventBridge
- Amazon DevOps Guru

- AWS Config
- AWS Trusted Advisor
- AWS CDK
- AWS Systems Manager
- Marketplace AWS

REL01-BP05 Automatizza la gestione delle quote

Implementa strumenti per ricevere avvisi quando le soglie stanno per essere raggiunte. Puoi automatizzare le richieste di aumento delle quote utilizzando AWS Service APIs Quotas.

Se integri il tuo Configuration Management Database (CMDB) o il sistema di ticketing con Service Quotas, puoi automatizzare il monitoraggio delle richieste di aumento delle quote e delle quote correnti. Oltre a AWS SDK, Service Quotas offre l'automazione utilizzando AWS Command Line Interface ()AWS CLI.

Anti-pattern comuni:

- Monitoraggio di quote e nei fogli di calcolo.
- Predisposizione di report sull'utilizzo giornaliero, settimanale o mensile e successivo confronto dell'utilizzo con le quote.

Vantaggi derivanti dall'adozione di questa best practice: il monitoraggio automatico delle quote di AWS servizio e il monitoraggio dell'utilizzo rispetto a tale quota consentono di vedere quando ci si avvicina a una quota. Puoi configurare l'automazione affinché ti aiuti a richiedere un aumento della quota quando necessario. Puoi decidere di ridurre alcune quote quando il tuo utilizzo tende alla direzione opposta per ottenere i vantaggi di riduzione del rischio (in caso di credenziali compromesse) e dei costi.

Livello di rischio associato se questa best practice non fosse adottata: medio

Guida all'implementazione

- Imposta il monitoraggio automatico SDKs Implementa strumenti che avvisano l'utente quando vengono raggiunte le soglie.
  - Utilizza Service Quotas e amplia il servizio con una soluzione di monitoraggio automatico delle quote, come AWS Limit Monitor o un'offerta di. Marketplace AWS

    - What is Service Quotas?
    - Quota Monitor attivo - Soluzione AWSAWS
- Imposta risposte automatiche basate su soglie di quota, utilizzando Amazon SNS e Service AWS Quotas. APIs
- Automazione dei test.
    - Configurazione delle soglie delle restrizioni.
    - Esegui l'integrazione con eventi di cambiamento AWS Config, pipeline di distribuzione EventBridge, Amazon o terze parti.
    - Impostazione artificiale di soglie basse per le quote in modo da testare le risposte.
    - Configurazione di operazioni automatizzate per eseguire azioni adeguate in seguito alle notifiche e contatta AWS Support se necessario.
    - Avvio manuale degli eventi di modifica.
    - Svolgimento di una giornata di gioco per testare il processo di modifica dell'aumento delle quote.

Risorse

Documenti correlati:

- APNPartner: partner che possono aiutarti nella gestione della configurazione
- Marketplace AWS: CMDB prodotti che aiutano a tenere traccia dei limiti
- AWS Service Quotas (precedentemente definite restrizioni dei servizi)
- AWS Trusted Advisor Controlli relativi alle migliori pratiche (consulta la sezione Limiti del servizio)
- Quota Monitor attivo AWS - AWS Soluzione
- Limiti EC2 del servizio Amazon
- What is Service Quotas?

Video correlati:

- AWS Live re:INforce 2019 - Service Quotas

REL01-BP06 Assicurarsi che esista un intervallo sufficiente tra le quote correnti e l'utilizzo massimo per consentire il failover

Il presente articolo illustra come mantenere lo spazio tra la quota di risorse e l'utilizzo e i relativi vantaggi per la tua organizzazione. Una volta terminato l'utilizzo di una risorsa, la quota di utilizzo può continuare a essere conteggiata per tale risorsa, con possibile conseguenza di una risorsa in errore o inaccessibile. Evita tale errore nelle risorse verificando che le quote tengano conto della sovrapposizione di risorse in errore o inaccessibili e della rispettiva sostituzione. Prendi in considerazioni casi come errori della rete, errori della zona di disponibilità o errori della regione durante il calcolo di questo divario.

Risultato desiderato: è possibile coprire piccoli o grandi errori nelle risorse o nell'accessibilità delle risorse entro le attuali soglie di servizio, tenendo conto degli errori delle zone, di rete o addirittura regionali nella pianificazione delle risorse.

Anti-pattern comuni:

- Impostazione delle quote di servizio in base alle esigenze attuali senza tenere conto degli scenari di failover.
- Calcolo della quota massima per un servizio senza tenere conto dei principali della stabilità statica.
- Calcolo della quota totale necessaria per ogni regione senza tenere conto delle potenziali risorse inaccessibili.
- Non sono stati presi in considerazione i limiti di isolamento dagli errori di AWS servizio per alcuni servizi e i loro potenziali modelli di utilizzo anomali.

Vantaggi dell'adozione di questa best practice: in caso di eventi di interruzione del servizio che influiscono sulla disponibilità dell'applicazione, utilizza il cloud per implementare strategie di ripristino da tali eventi. Un esempio di strategia consiste nella creazione di risorse aggiuntive per sostituire quelle inaccessibili e soddisfare le condizioni di failover senza esaurire il limite del servizio.

Livello di rischio associato se questa best practice non fosse adottata: medio

Guida all'implementazione

Nel valutare un limite di quota, tieni conto dei casi di failover che possono verificarsi a causa di un peggioramento della situazione. Considera i casi di failover seguenti:

- Un file interrotto o inaccessibile. VPC
- Sottorete inaccessibile.

- Zona di disponibilità degradata che influisce sull'accessibilità delle risorse.
- Diversi instradamenti di rete o punti di ingresso e uscita bloccati o modificati.
- Impatto di una regione degradata sull'accessibilità delle risorse.
- Errore in un sottoinsieme di risorse in una regione o in una zona di disponibilità.

La decisione relativa all'avvio del failover è unica per ogni situazione, in quanto l'impatto aziendale può variare. Gestisci la pianificazione della capacità delle risorse nella posizione di failover e le quote delle risorse prima di decidere di effettuare il failover di un'applicazione o di un servizio.

Prendi in considerazione i picchi di attività più elevati del normale nell'esame delle quote per ciascun servizio. Questi picchi potrebbero essere correlati a risorse ancora attive ma inaccessibili a causa di reti o autorizzazioni. Le risorse attive non terminate vengono conteggiate rispetto al limite di quota del servizio.

Passaggi dell'implementazione

- Mantieni uno spazio sufficiente tra la quota di servizio e l'utilizzo massimo in modo da gestire un failover o la perdita di accessibilità.
- Determina le quote di servizio. Tieni conto di modelli di implementazione tipici, requisiti di disponibilità e crescita dei consumi.
- Richiedi aumenti delle quote, se necessario. Prevedi un tempo di attesa per la richiesta di aumento della quota.
- Determina i requisiti di affidabilità, noti anche come numero di 9.
- Analizza i potenziali scenari di errore, come la perdita di un componente, di una zona di disponibilità o di una regione.
- Stabilisci la metodologia di implementazione (ad esempio, canary, blu/verde, rosso/nero e rolling).
- Includi un buffer appropriato rispetto al limite della quota attuale. Un esempio di buffer potrebbe essere del 15%.
- Includi calcoli per la stabilità statica (zonale e regionale) laddove appropriato.
- Pianifica la crescita dei consumi e monitora i trend di consumo.
- Tieni conto dell'impatto della stabilità statica per i carichi di lavoro più critici. Valuta la conformità delle risorse a un sistema statisticamente stabile in tutte le regioni e le zone di disponibilità.
- Valuta l'utilizzo di prenotazioni della capacità on demand per pianificare la capacità in anticipo rispetto a qualsiasi failover. Si tratta di una strategia utile da implementare per le pianificazioni

aziendali critiche per ridurre i possibili rischi legati all'ottenimento della quantità e del tipo di risorse corretti durante il failover.

Risorse

Best practice correlate:

- REL01-BP01 Conoscenza delle quote e dei vincoli di servizio
- REL01-BP02 Gestisci le quote di servizio tra account e regioni
- REL01-BP03 Soddisfa quote e vincoli di servizio fissi tramite l'architettura
- REL01-BP04 Monitoraggio e gestione delle quote
- REL01-BP05 Automatizza la gestione delle quote
- REL03-BP01 Scegli come segmentare il tuo carico di lavoro
- REL10-BP01 Implementazione del carico di lavoro in più sedi
- REL11-BP01 Monitora tutti i componenti del carico di lavoro per rilevare i guasti
- REL11-BP03 Automatizza la guarigione su tutti i livelli
- REL12-BP05 Verifica la resilienza utilizzando l'ingegneria del caos

Documenti correlati:

- AWS Il pilastro dell'affidabilità di Well-Architected Framework: disponibilità
- AWS Service Quotas (precedentemente denominate limiti di servizio)
- AWS Trusted Advisor Controlli relativi alle migliori pratiche (vedere la sezione Limiti del servizio)
- AWS limita il monitoraggio delle AWS risposte
- Limiti EC2 del servizio Amazon
- What is Service Quotas?
- How to Request Quota Increase
- Service endpoints and quotas
- Guida per l'utente di Service Quotas
- Quota Monitor per AWS
- AWS Limiti di isolamento dei guasti
- Availability with redundancy

- AWS per i dati
- Cos'è l'integrazione continua?
- Cos'è la distribuzione continua?
- APNPartner: partner che possono aiutare nella gestione della configurazione
- Gestione del ciclo di vita dell'account in ambienti SaaS account-per-tenant su AWS
- Gestione e monitoraggio della limitazione dei carichi di API lavoro
- Visualizza i AWS Trusted Advisor consigli su larga scala con AWS Organizations
- Automatizzazione degli aumenti dei limiti di servizio e del supporto aziendale con AWS Control Tower
- Actions, resources, and condition keys for Service Quotas

Video correlati:

- AWS Live re:INforce 2019 - Service Quotas
- Visualizzazione e gestione delle quote per AWS i servizi tramite Service Quotas
- AWS IAMDimostrazione di Quotas
- AWS re:Invent 2018: Circuiti chiusi e menti aperte: come assumere il controllo di sistemi, grandi e piccoli

Strumenti correlati:

- AWS CodeDeploy
- AWS CloudTrail
- Amazon CloudWatch
- Amazon EventBridge
- Amazon DevOps Guru
- AWS Config
- AWS Trusted Advisor
- AWS CDK
- AWS Systems Manager
- Marketplace AWS

REL2. Come si pianifica la topologia della rete?

I carichi di lavoro sono spesso presenti in più ambienti. Questi includono più ambienti cloud (sia accessibili pubblicamente sia privati) e, possibilmente, l'infrastruttura del data center esistente. I piani devono includere considerazioni di rete, ad esempio connettività intrasistema e intersistema, gestione di indirizzi IP pubblici, gestione di indirizzi IP privati e risoluzione dei nomi di dominio.

Best practice

- REL02-BP01 Utilizza una connettività di rete ad alta disponibilità per gli endpoint pubblici del carico di lavoro
- REL02-BP02 Fornisci connettività ridondante tra reti private nel cloud e ambienti on-premise
- REL02-BP03 Garantire che l'allocazione della sottorete IP tenga conto dell'espansione e della disponibilità
- REL02-BP04 Preferisce le topologie alle mesh hub-and-spoke many-to-many
- REL02-BP05 Applica intervalli di indirizzi IP privati non sovrapposti in tutti gli spazi di indirizzi privati a cui sono connessi

REL02-BP01 Utilizza una connettività di rete ad alta disponibilità per gli endpoint pubblici del carico di lavoro

La creazione di una connettività di rete ad alta disponibilità verso gli endpoint pubblici dei carichi di lavoro può aiutarvi a ridurre i tempi di inattività dovuti alla perdita di connettività e a migliorare la disponibilità e il carico di lavoro. SLA A tal fine, utilizzate reti di distribuzione dei contenuti (CDNs)DNS, API gateway, sistemi di bilanciamento del carico o proxy inversi ad alta disponibilità.

Risultato desiderato: la pianificazione, la realizzazione e la messa in funzione di una connettività di rete altamente disponibile per i tuoi endpoint pubblici è fondamentale. Se il carico di lavoro diventa irraggiungibile a causa della perdita di connettività, il sistema apparirà ai clienti come non funzionante, anche se il carico di lavoro è in esecuzione e disponibile. Combinando connettività di rete a disponibilità elevata e resiliente per gli endpoint pubblici del carico di lavoro, a un'architettura resiliente per il carico di lavoro stesso, puoi offrire ai clienti la disponibilità e il livello di servizio migliori possibili.

AWS Global Accelerator, Amazon CloudFront, Amazon API Gateway URLs AWS AppSync APIs, AWS Lambda Function ed Elastic Load Balancing (ELB) forniscono tutti endpoint pubblici ad alta disponibilità. Amazon Route 53 fornisce un DNS servizio ad alta disponibilità per la risoluzione dei nomi di dominio per verificare che gli indirizzi degli endpoint pubblici possano essere risolti.

Puoi anche valutare dispositivi Marketplace AWS software per il bilanciamento del carico e il proxy.

Anti-pattern comuni:

- Progettazione di un carico di lavoro ad alta disponibilità senza pianificazione DNS e connettività di rete per un'elevata disponibilità.
- Utilizzo di indirizzi Internet pubblici su singole istanze o contenitori e gestione della connettività con essi. DNS
- Uso di indirizzi IP anziché nomi di dominio per l'individuazione dei servizi.
- Mancata esecuzione di test su scenari con perdita di connettività agli endpoint pubblici.
- Mancata analisi delle esigenze di throughput della rete e dei modelli di distribuzione.
- Nessuna attività di test e pianificazione per scenari di possibile interruzione della connettività di rete Internet agli endpoint pubblici del carico di lavoro.
- Distribuzione di contenuti (pagine Web, asset statici o file multimediali) in un'area geografica di grandi dimensioni senza l'uso di una rete di distribuzione di contenuti.
- Nessuna pianificazione di attacchi Distributed Denial of Service (DDoS). DDoSgli attacchi rischiano di bloccare il traffico legittimo e ridurre la disponibilità per gli utenti.

Vantaggi dell'adozione di questa best practice: la progettazione pensata per una connettività di rete a elevata disponibilità e resilienza garantisce l'accessibilità e la disponibilità del carico di lavoro agli utenti.

Livello di rischio associato se questa best practice non fosse adottata: elevato

Guida all'implementazione

Alla base della creazione di connettività di rete a disponibilità elevata agli endpoint pubblici vi è l'instradamento del traffico. Per verificare che il traffico sia in grado di raggiungere gli endpoint, questi DNS devono essere in grado di risolvere i nomi di dominio negli indirizzi IP corrispondenti. Utilizza un Domain Name System ad alta disponibilità e scalabilità (DNS) come Amazon Route 53 per gestire i DNS record del tuo dominio. Puoi usare anche i controlli dell'integrità forniti da Amazon Route 53. I controlli di integrità verificano che l'applicazione sia raggiungibile, disponibile e funzionale e possono essere configurati in modo da imitare il comportamento dell'utente, ad esempio richiedendo una pagina Web o una pagina specifica. URL In caso di errore, Amazon Route 53 risponde alle richieste di DNS risoluzione e indirizza il traffico solo verso endpoint integri. Puoi anche prendere in considerazione l'utilizzo delle funzionalità di routing basato sulla geolocalizzazione DNS e sulla latenza offerte da Amazon Route 53.

Per verificare che il carico di lavoro stesso sia altamente disponibile, usa Elastic Load Balancing ELB (). Amazon Route 53 può essere utilizzato per indirizzare il traffico versoELB, che distribuisce il traffico alle istanze di elaborazione di destinazione. Puoi anche utilizzare Amazon API Gateway insieme AWS Lambda a una soluzione serverless. I clienti possono anche eseguire carichi di lavoro multipli. Regioni AWS Grazie a un pattern attivo/attivo multisito, il carico di lavoro può servire il traffico proveniente da più regioni. Con un pattern attivo/passivo multisito, il carico di lavoro serve il traffico proveniente dalla regione attiva, mentre nella regione secondaria avviene la replica dei dati, che diventano attivi in caso di guasto nella regione primaria. I controlli dello stato di Route 53 possono quindi essere utilizzati per controllare il DNS failover da qualsiasi endpoint in una regione primaria a un endpoint in una regione secondaria, verificando che il carico di lavoro sia raggiungibile e disponibile per gli utenti.

Amazon CloudFront offre una soluzione semplice API per la distribuzione di contenuti con bassa latenza e velocità di trasferimento dati elevate soddisfacendo le richieste utilizzando una rete di edge location in tutto il mondo. Le reti di distribuzione dei contenuti (CDNs) servono i clienti offrendo contenuti localizzati o memorizzati nella cache in una posizione vicina all'utente. Ciò migliora anche la disponibilità dell'applicazione poiché il carico dei contenuti viene spostato dai server alle sedi CloudFront periferiche. Le posizioni edge e le cache edge regionali includono copie memorizzate nella cache del contenuto vicino agli utenti, per il recupero rapido e una raggiungibilità e una disponibilità maggiori del carico di lavoro.

Per carichi di lavoro con utenti distribuiti geograficamente, AWS Global Accelerator consente di migliorare la disponibilità e le prestazioni delle applicazioni. AWS Global Accelerator fornisce indirizzi IP statici Anycast che fungono da punto di accesso fisso all'applicazione ospitata in una o più applicazioni. Regioni AWS Ciò consente al traffico di entrare nella rete AWS globale il più vicino possibile agli utenti, migliorando la raggiungibilità e la disponibilità del carico di lavoro. AWS Global Accelerator monitora inoltre lo stato degli endpoint delle applicazioni utilizzando e controllando lo TCP stato di integrità. HTTP HTTPS Eventuali variazioni dell'integrità o della configurazione degli endpoint permettono il reindirizzamento del traffico degli utenti a endpoint integri che offrono le prestazioni e la disponibilità migliori agli utenti. Inoltre, AWS Global Accelerator dispone di un design di isolamento dei guasti che utilizza due IPv4 indirizzi statici serviti da zone di rete indipendenti che aumentano la disponibilità delle applicazioni.

Per aiutare a proteggere i clienti dagli attacchi, fornisce. DDoS AWS AWS Shield Standard Shield Standard si attiva automaticamente e protegge dagli attacchi delle infrastrutture comuni (layer 3 e 4) come SYN UDP /floods e attacchi di riflessione per supportare l'elevata disponibilità delle applicazioni. AWS Per ulteriori protezioni contro attacchi più sofisticati e più estesi (come le UDP inondazioni), attacchi di esaurimento dello stato (come le TCP SYN inondazioni) e per proteggere le

applicazioni in esecuzione su Amazon Elastic Compute Cloud (EC2Amazon), Elastic Load Balancing (ELB), CloudFront Amazon AWS Global Accelerator e Route 53, puoi prendere in considerazione l'utilizzo. AWS Shield Advanced Per proteggerti da attacchi a livello di applicazione come o flood, usa. HTTP POST GET AWS WAF AWS WAF può utilizzare indirizzi IP, HTTP intestazioni, HTTP body, URI stringhe, SQL injection e condizioni di cross-site scripting per determinare se una richiesta deve essere bloccata o consentita.

Passaggi dell'implementazione

1. Configurazione a disponibilità elevataDNS: Amazon Route 53 è un servizio Web di Domain Name System (DNS) ad alta disponibilità e scalabilità. Route 53 collega le richieste degli utenti alle applicazioni Internet in esecuzione in locale AWS o in locale. Per ulteriori informazioni, consulta la sezione Configurazione di Amazon Route 53 come DNS servizio.
2. Configura controlli dell'integrità: quando usi Route 53, verifica che solo le destinazioni integre siano risolvibili. Inizia creando i controlli di integrità di Route 53 e configurando DNS il failover. Nel configurare controlli dell'integrità, è importante tenere conto degli aspetti seguenti:
   a. Modo in cui Amazon Route 53 determina se un controllo dell'integrità ha esito positivo
   b. Creazione, aggiornamento ed eliminazione di controlli dell'integrità
   c. Monitoraggio dello stato dei controlli dell'integrità e ricezione di notifiche
   d. Le migliori pratiche per Amazon Route 53 DNS
3. Connect il tuo DNS servizio ai tuoi endpoint.
   a. In caso di utilizzo di Elastic Load Balancing come target per il tuo traffico, crea un record di alias mediante Amazon Route 53 che punti all'endpoint regionale del tuo sistema bilanciatore del carico. Durante la creazione del record di alias, imposta l'opzione Valutazione dello stato target su Sì.
   b. Per carichi di lavoro serverless o privati APIs quando si utilizza API Gateway, utilizza Route 53 per indirizzare il traffico verso Gateway. API
4. Opta per una rete di distribuzione di contenuti (CDN).
   a. Per distribuire contenuti utilizzando postazioni periferiche più vicine all'utente, inizia con il comprendere in che modo vengono CloudFront distribuiti i contenuti.
   b. Inizia con una CloudFront distribuzione semplice. CloudFront quindi sa da dove desideri che vengano distribuiti i contenuti e i dettagli su come monitorare e gestire la distribuzione dei contenuti. I seguenti aspetti sono importanti da comprendere e considerare quando si imposta CloudFront la distribuzione:
      i. Come funziona la memorizzazione nella cache con le CloudFront edge location

   ii. Aumento della percentuale di richieste che vengono servite direttamente dalle CloudFront cache (rapporto di successo della cache)
   iii. Utilizzo di Amazon CloudFront Origin Shield
   iv. Ottimizzazione dell'alta disponibilità con il failover di CloudFront origine
5. Imposta la protezione a livello di applicazione: ti AWS WAF aiuta a proteggerti da exploit e bot Web comuni che possono influire sulla disponibilità, compromettere la sicurezza o consumare risorse eccessive. Per una comprensione più approfondita, scopri come AWS WAF funziona e quando sei pronto per implementare le protezioni dalle HTTP POST AND GET inondazioni a livello applicativo, consulta la sezione Guida introduttiva. AWS WAF Puoi anche utilizzare AWS WAF con CloudFront consulta la documentazione su come AWS WAF funziona con CloudFront le funzionalità di Amazon.
6. Imposta una DDoS protezione aggiuntiva: per impostazione predefinita, tutti AWS i clienti ricevono protezione dagli DDoS attacchi più comuni e più frequenti a livello di rete e trasporto che prendono di mira il tuo sito Web o la tua applicazione senza AWS Shield Standard costi aggiuntivi. Per una protezione aggiuntiva delle applicazioni con accesso a Internet in esecuzione su AmazonEC2, Elastic Load Balancing, Amazon e Amazon Route 53 CloudFront AWS Global Accelerator, puoi AWS Shield Advancedprendere in considerazione ed esaminare esempi di architetture resilienti. DDoS Per proteggere il carico di lavoro e gli endpoint pubblici dagli attacchi, consulta la sezione Guida introduttiva. DDoS AWS Shield Advanced

Risorse

Best practice correlate:

- REL10-BP01 Implementazione del carico di lavoro in più sedi
- REL10-BP02 Seleziona le posizioni appropriate per l'implementazione in più sedi
- REL11-BP04 Affidati al piano dati e non al piano di controllo durante il ripristino
- REL11-BP06 Invia notifiche quando gli eventi influiscono sulla disponibilità

Documenti correlati:

- APNPartner: partner che possono aiutarti a pianificare la tua rete
- Marketplace AWS per l'infrastruttura di rete
- Che cos'è AWS Global Accelerator?
- Che cos'è Amazon CloudFront?

- What is Amazon Route 53?
- Cos'è l'Elastic Load Balancing?
- Network Connectivity capability - Establishing Your Cloud Foundations
- Che cos'è Amazon API Gateway?
- Cosa sono AWS WAFAWS Shield, e AWS Firewall Manager?
- Cos'è Amazon Application Recovery Controller?
- Configura controlli sanitari personalizzati per il DNS failover

Video correlati:

- AWS re:Invent 2022 - Migliora le prestazioni e la disponibilità con AWS Global Accelerator
- AWS re:Invent 2020: gestione globale del traffico con Amazon Route 53
- AWS re:Invent 2022 - Utilizzo di applicazioni Multi-AZ ad alta disponibilità
- AWS re:Invent 2022 - Approfondisci l'infrastruttura di rete AWS
- AWS re:Invent 2022 - Costruire reti resilienti

Esempi correlati:

- Ripristino di emergenza con Amazon Application Recovery Controller (ARC)
- Workshop sull'affidabilità
- AWS Global Accelerator Workshop

REL02-BP02 Fornisci connettività ridondante tra reti private nel cloud e ambienti on-premise

Implementa la ridondanza delle connessioni tra reti private nel cloud e negli ambienti on-premises per ottenere la resilienza della connettività. A tal fine, puoi implementare due o più collegamenti e percorsi di traffico, preservando la connettività in caso di errori di rete.

Anti-pattern comuni:

- Dipendi da una sola connessione di rete, che crea un singolo punto di errore.
- Si utilizza solo uno o più VPN tunnel che terminano nella stessa zona di disponibilità.
- Ti affidi a uno ISP per la VPN connettività, che può portare a guasti completi durante ISP le interruzioni.

- Non implementate protocolli di routing dinamici come quelli BGP fondamentali per reindirizzare il traffico durante le interruzioni della rete.
- Ignorate i limiti di larghezza di banda dei VPN tunnel e sopravvalutate le loro capacità di backup.

Vantaggi dell'adozione di questa best practice: implementando una connettività ridondante tra il tuo ambiente cloud e l'ambiente aziendale oppure on-premises, puoi garantire l'affidabilità delle comunicazioni dei servizi dipendenti tra due ambienti.

Livello di rischio associato se questa best practice non fosse adottata: elevato

Guida all'implementazione

Quando si utilizza AWS Direct Connect per connettere la rete locale a AWS, è possibile ottenere la massima resilienza della rete (SLAdel 99,99%) utilizzando connessioni separate che terminano su dispositivi distinti in più di una posizione locale e più di una posizione. AWS Direct Connect Questa topologia offre resilienza ai guasti dei dispositivi, ai problemi di connettività e alle interruzioni complete della posizione. In alternativa, è possibile ottenere un'elevata resilienza (SLAdel 99,9%) utilizzando due connessioni individuali a più sedi (ciascuna sede locale connessa a un'unica posizione Direct Connect). Questo approccio offre protezione dalle interruzioni della connettività causate da interruzioni della fibra o guasti dei dispositivi e aiuta a mitigare le interruzioni complete della posizione. Il AWS Direct Connect Resiliency Toolkit può aiutarti a progettare la topologia. AWS Direct Connect

Puoi anche prendere in considerazione l'idea di AWS Site-to-Site VPN terminare con un backup economico sulla tua connessione principale. AWS Transit Gateway AWS Direct Connect Questa configurazione consente il routing multipath (ECMP) a parità di costo su più VPN tunnel, con una velocità di trasmissione fino a 50 Gbps, anche se ogni tunnel è limitato a 1,25 Gbps. VPN È importante notare, tuttavia, che questa è ancora la scelta più efficace per ridurre al minimo le interruzioni di rete e AWS Direct Connect fornire una connettività stabile.

Quando utilizzi VPNs Internet per connettere l'ambiente cloud al data center locale, configura due VPN tunnel come parte di un'unica connessione. site-to-site VPN Ogni tunnel deve terminare in una zona di disponibilità diversa per garantire l'alta disponibilità e utilizzare hardware ridondante per prevenire gli errori dei dispositivi on-premises. Inoltre, prendi in considerazione più connessioni Internet da vari provider di servizi Internet (ISPs) presso la tua sede locale per evitare interruzioni complete della VPN connettività dovute a una singola interruzione. ISP La scelta ISPs con routing e infrastrutture diverse, in particolare quelle con percorsi fisici separati verso gli AWS endpoint, offre un'elevata disponibilità di connettività.

Oltre alla ridondanza fisica con più AWS Direct Connect connessioni e più VPN tunnel (o una combinazione di entrambi), è fondamentale anche l'implementazione del routing dinamico del Border Gateway Protocol (BGP). Dynamic BGP fornisce il reindirizzamento automatico del traffico da un percorso all'altro in base alle condizioni di rete in tempo reale e alle politiche configurate. Questo comportamento dinamico è particolarmente utile per mantenere la disponibilità della rete e la continuità del servizio in caso di errori di collegamento o rete. Seleziona rapidamente percorsi alternativi, migliorando la resilienza e l'affidabilità della rete.

Passaggi dell'implementazione

- Acquisizione di connettività ad alta disponibilità tra e AWS l'ambiente locale.
  - Utilizza più AWS Direct Connect connessioni o VPN tunnel tra reti private distribuite separatamente.
  - Utilizza più AWS Direct Connect sedi per un'elevata disponibilità.
  - Se ne utilizzi più Regioni AWS, crea ridondanza in almeno due di esse.
- AWS Transit GatewayUtilizzatelo, quando possibile, per terminare la VPN connessione.
- Valuta i Marketplace AWS dispositivi a cui terminare VPNs o estendere la WAN porta SD- AWS. Se utilizzi Marketplace AWS appliance, implementa istanze ridondanti per un'elevata disponibilità in diverse zone di disponibilità.
- Fornisci una connessione ridondante all'ambiente on-premises.
  - Potresti aver bisogno di connessioni ridondanti a più connessioni per soddisfare le tue esigenze di disponibilità Regioni AWS .
  - Usa l'AWS Direct Connect Resiliency Toolkit per iniziare.

Risorse

Documenti correlati:

- AWS Direct Connect Raccomandazioni sulla resilienza
- Utilizzo di Site-to-Site VPN connessioni ridondanti per fornire il failover
- Politiche e comunità di routing BGP
- Configurazioni attive/attive e attive/passive in AWS Direct Connect
- APNPartner: partner che possono aiutarti a pianificare la tua rete
- Marketplace AWS per l'infrastruttura di rete
- Amazon Virtual Private Cloud Connectivity Options Whitepaper

- Creazione di un'infrastruttura multirete scalabile e sicura VPC AWS
- Utilizzo di connessioni ridondanti Site-to-Site VPN per fornire il failover
- Utilizzo del AWS Direct Connect Resiliency Toolkit per iniziare
- VPCEndpoint e servizi per gli VPC endpoint ()AWS PrivateLink
- Che cos'è AmazonVPC?
- What is a transit gateway?
- Che cos'è AWS Site-to-Site VPN?
- Working with Direct Connect gateways

Video correlati:

- AWS re:Invent 2018: VPC design avanzato e nuove funzionalità per Amazon VPC
- AWS re:Invent 2019: architetture di riferimento per molti AWS Transit Gateway VPCs

REL02-BP03 Garantire che l'allocazione della sottorete IP tenga conto dell'espansione e della disponibilità

Gli intervalli di indirizzi VPC IP di Amazon devono essere sufficientemente ampi da soddisfare i requisiti dei carichi di lavoro, incluso il calcolo delle future espansioni e dell'allocazione degli indirizzi IP alle sottoreti nelle zone di disponibilità. Ciò include sistemi di bilanciamento del carico, istanze e applicazioni basate su contenitori. EC2

Quando si pianifica la topologia di rete, il primo passo è definire lo spazio stesso degli indirizzi IP. Gli intervalli di indirizzi IP privati (secondo le linee guida del RFC 1918) devono essere assegnati per ciascuno. VPC Nell'ambito di questo processo, soddisfa i seguenti requisiti:

- Consenti lo spazio degli indirizzi IP per più di uno VPC per regione.
- All'interno di aVPC, consenti spazio per più sottoreti in modo da poter coprire più zone di disponibilità.
- Valuta la possibilità di lasciare dello spazio inutilizzato nei CIDR blocchi VPC per future espansioni.
- Assicurati che sia disponibile uno spazio di indirizzi IP per soddisfare le esigenze di eventuali flotte transitorie di EC2 istanze Amazon che potresti utilizzare, ad esempio flotte Spot per l'apprendimento automatico, cluster Amazon o cluster Amazon EMR Redshift. Un'analoga considerazione dovrebbe essere data ai cluster Kubernetes, come Amazon Elastic Kubernetes

Service (EKSAmazon), poiché a ogni pod Kubernetes viene assegnato un indirizzo routabile dal blocco per impostazione predefinita. VPC CIDR

- Tieni presente che i primi quattro indirizzi IP e l'ultimo indirizzo IP in ogni blocco di sottorete sono riservati e non disponibili per l'uso. CIDR
- Tieni presente che il VPC CIDR blocco iniziale assegnato al tuo VPC non può essere modificato o eliminato, ma puoi aggiungere ulteriori blocchi non sovrapposti CIDR a. VPC La sottorete IPv4 CIDRs non può essere modificata, tuttavia sì. IPv6 CIDRs
- Il VPC CIDR blocco più grande possibile è /16 e il più piccolo è /28.
- Prendi in considerazione altre reti connesse (VPClocali o altri provider di servizi cloud) e assicurati che lo spazio degli indirizzi IP non si sovrapponga. Per ulteriori informazioni, consulta REL02-BP05 Applica intervalli di indirizzi IP privati non sovrapposti in tutti gli spazi di indirizzi privati a cui sono connessi.

Risultato desiderato: una sottorete IP scalabile può aiutarti a far fronte alla crescita futura e a evitare inutili sprechi.

Anti-pattern comuni:

- Non considerare la crescita futura, con conseguenti CIDR blocchi troppo piccoli che richiedono una riconfigurazione, con conseguenti potenziali tempi di inattività.
- Stima erronea del numero di indirizzi IP utilizzabili da un bilanciatore del carico elastico.
- Distribuzione di numerosi bilanciatori del carico a traffico elevato nelle stesse sottoreti.
- Utilizzo di meccanismi di dimensionamento automatico senza monitorare il consumo di indirizzi IP.
- Definire CIDR intervalli eccessivamente ampi ben oltre le aspettative di crescita future, il che può comportare difficoltà nel peering con altre reti con intervalli di indirizzi sovrapposti.

Vantaggi dell'adozione di questa best practice: in questo modo puoi consentire la crescita dei carichi di lavoro e continuare a fornire disponibilità nell'aumentare verticalmente.

Livello di rischio associato se questa best practice non fosse adottata: medio

Guida all'implementazione

Pianifica la tua rete in base a crescita, compliance normativa e integrazione con altre reti. Senza una pianificazione adeguata, la crescita può essere sottovalutata, la compliance normativa può cambiare e l'implementazione di acquisizioni o di connessioni a reti private può rivelarsi difficile.

- Seleziona le aree Account AWS e le regioni pertinenti in base ai requisiti di servizio, di latenza, normativi e di disaster recovery (DR).
- Identifica le tue esigenze per le VPC implementazioni regionali.
- Identifica la dimensione di. VPCs
  - Determina se intendi implementare la connettività multipla. VPC
    - What Is a Transit Gateway?
    - Connettività multipla a regione singola VPC
  - Stabilisci se hai bisogno della segregazione delle reti a causa di requisiti normativi.
  - Crea VPCs con CIDR blocchi di dimensioni adeguate per soddisfare le tue esigenze attuali e future.
    - Se avete proiezioni di crescita sconosciute, potreste optare per CIDR blocchi più grandi per ridurre il potenziale di future riconfigurazioni.
  - Prendi in considerazione l'utilizzo dell'IPv6indirizzamento per le sottoreti come parte di un dual-stack. VPC IPv6è ideale per essere utilizzato in sottoreti private contenenti flotte di istanze o contenitori temporanei che altrimenti richiederebbero un gran numero di indirizzi. IPv4

Risorse

Best practice Well-Architected correlate:

- REL02-BP05 Applica intervalli di indirizzi IP privati non sovrapposti in tutti gli spazi di indirizzi privati a cui sono connessi

Documenti correlati:

- APNPartner: partner che possono aiutarti a pianificare la tua rete
- Marketplace AWS per l'infrastruttura di rete
- Amazon Virtual Private Cloud Connectivity Options Whitepaper
- Multiple data center HA network connectivity
- Connettività multipla a regione singola VPC
- Che cos'è AmazonVPC?
- IPv6su AWS
- IPv6sulle architetture di riferimento

- Amazon Elastic Kubernetes Service lancia il supporto IPv6
- Consigli per i tuoi sistemi Classic Load Balancer VPC
- Subnet delle zone di disponibilità - Application Load Balancer
- Zone di disponibilità - Network Load Balancer

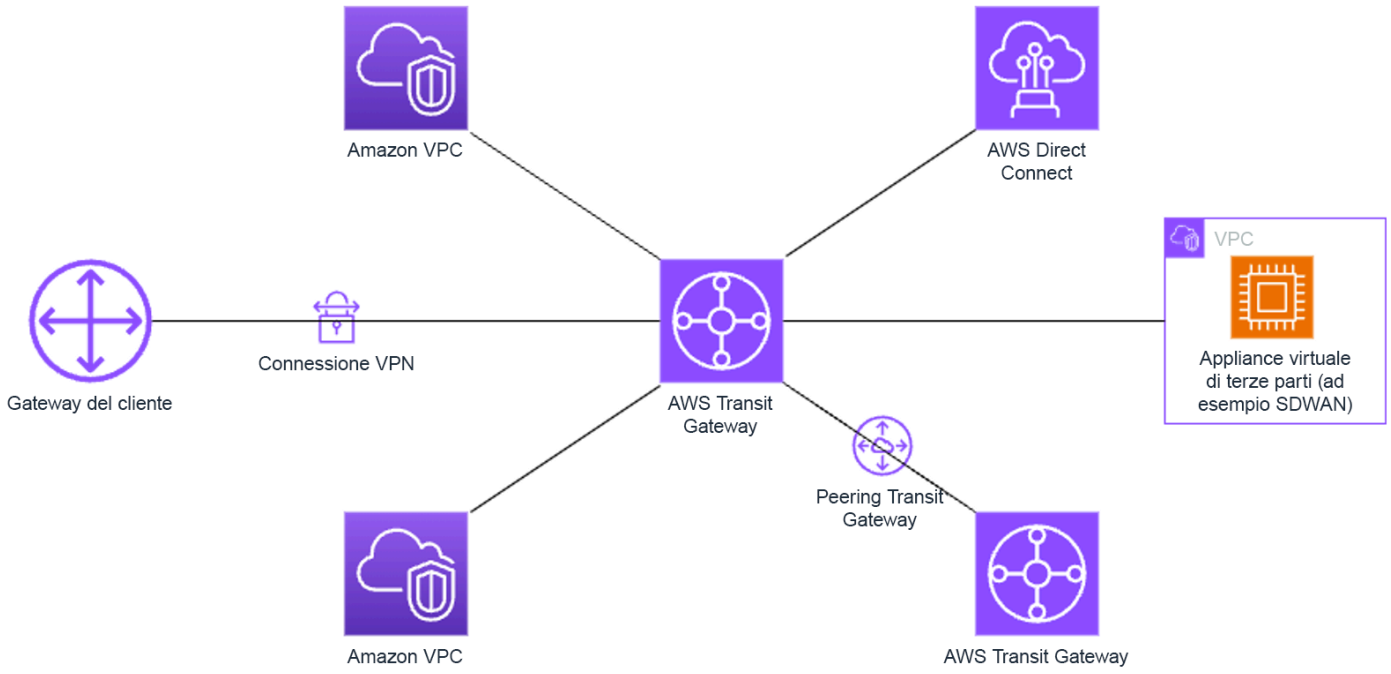
Video correlati:

- AWS re:Invent 2018: VPC design avanzato e nuove funzionalità per Amazon VPC (03) NET3
- AWS re:Invent 2019: architetture di AWS Transit Gateway riferimento per molti (06-R1) VPCs NET4
- AWS re:Invent 2023: pronti per il futuro? AWS Progettare reti per la crescita e la flessibilità (0) NET31

REL02-BP04 Preferisce le topologie alle mesh hub-and-spoke many-to-many

Quando connettete più reti private, come Virtual Private Clouds (VPCs) e reti locali, optate per una topologia anziché una hub-and-spoke rete mesh. A differenza delle topologie mesh, in cui ogni rete si connette direttamente alle altre e aumenta la complessità e il sovraccarico di gestione, l' hub-and-spokearchitettura centralizza le connessioni tramite un singolo hub. Questa centralizzazione semplifica la struttura della rete e ne migliora il funzionamento, la scalabilità e il controllo.

AWS Transit Gateway è un servizio gestito, scalabile e ad alta disponibilità progettato per la costruzione di reti su. hub-and-spoke AWS Funge da hub centrale della rete che fornisce la segmentazione, il routing centralizzato e la connessione semplificata agli ambienti cloud e on-premises. La figura seguente illustra come è possibile utilizzare AWS Transit Gateway per creare la topologia. hub-and-spoke



Anti-pattern comuni:

- Si complicano eccessivamente le politiche di routing in un' hub-and-spokearchitettura, il che riduce l'efficienza della rete e complica sia la risoluzione dei problemi che la gestione proattiva.
- Una segmentazione insufficiente basata sul routing all'interno dell'hub potrebbe comportare vulnerabilità che potenzialmente espongono la rete ad accessi non autorizzati.
- Senza un'attenta ottimizzazione, il traffico instradato attraverso l'hub può comportare elevati costi di trasferimento dei dati, in particolare per il traffico che attraversa zone di disponibilità e regioni. L'uso di efficaci strategie di gestione del traffico è essenziale per controllare le spese.

Vantaggi derivanti dall'adozione di questa best practice: con l'aumento del numero di reti connesse, la gestione e l'espansione della connettività mesh diventano sempre più impegnative. AWS Transit Gateway offre un hub gestito scalabile e affidabile per la costruzione e il funzionamento delle vostre hub-and-spoke topologie. Quando si utilizza AWS Transit Gateway, è possibile stabilire connessioni e centralizzare il routing del traffico su più reti.

Livello di rischio associato se questa best practice non fosse adottata: medio

Guida all'implementazione

- Pianifica la rete.

- Crea il tuo. AWS Transit Gateway
- Allega il tuoVPCs.
- Se necessario, create VPN connessioni o gateway Direct Connect e associateli al Transit Gateway.
- Definisci come viene instradato il traffico tra le connessioni VPCs connesse e le altre connessioni tramite la configurazione delle tabelle di routing del Transit Gateway.
- Usa Amazon CloudWatch per monitorare e modificare le configurazioni secondo necessità per l'ottimizzazione delle prestazioni e dei costi.

Risorse

Documenti correlati:

- What Is a Transit Gateway?
- Creazione di un'infrastruttura multirete scalabile e sicura VPC AWS
- Creazione di una rete globale utilizzando AWS Transit Gateway il peering interregionale
- Opzioni di connettività di Amazon Virtual Private Cloud
- APNPartner: partner che possono aiutarti a pianificare la tua rete
- Marketplace AWS per l'infrastruttura di rete

Video correlati:

- AWS re:Invent 2023 - fondamenti per il networking AWS
- AWS re:Invent 2023 - Design avanzati e nuove funzionalità VPC

REL02-BP05 Applica intervalli di indirizzi IP privati non sovrapposti in tutti gli spazi di indirizzi privati a cui sono connessi

Gli intervalli di indirizzi IP di ciascuno di voi non VPCs devono sovrapporsi in caso di peering, connessione tramite Transit Gateway o connessione tramite rete. VPN Evita i conflitti di indirizzi IP tra ambienti a VPC e locali o con altri provider di servizi cloud che utilizzi. Bisogna inoltre disporre di una soluzione per allocare gli intervalli di indirizzi IP privati quando necessario. Un sistema di gestione degli indirizzi IP (IPAM) può aiutare ad automatizzare questa operazione.

Risultato desiderato:

- Nessun conflitto di intervalli di indirizzi IP tra VPCs ambienti locali o altri provider di servizi cloud.

- La corretta gestione degli indirizzi IP consente di scalare più facilmente l'infrastruttura di rete per supportare la crescita e i cambiamenti dei requisiti di rete.

Anti-pattern comuni:

- Utilizzate lo stesso intervallo di IP che avete VPC in sede, nella rete aziendale o in altri provider di servizi cloud
- Non si tiene traccia degli intervalli di IP VPCs utilizzati per distribuire i carichi di lavoro.
- Ricorso a processi manuali di gestione degli indirizzi IP, come i fogli di calcolo.
- CIDRBlocchi sovradimensionati o sottodimensionati, con conseguente spreco di indirizzi IP o spazio di indirizzamento insufficiente per il carico di lavoro.

Vantaggi dell'adozione di questa best practice: la pianificazione attiva della rete garantisce di non avere più occorrenze dello stesso indirizzo IP nelle reti interconnesse. In questo modo si evitano problemi di instradamento in parti del carico di lavoro che utilizzano le diverse applicazioni.

Livello di rischio associato se questa best practice non fosse adottata: medio

Guida all'implementazione

Utilizza un programmaIPAM, come Amazon VPC IP Address Manager, per monitorare e gestire il tuo CIDR utilizzo. Diversi IPAMs sono disponibili anche presso Marketplace AWS. Valuta il tuo potenziale utilizzo AWS, aggiungi CIDR intervalli a quelli esistenti VPCs e crea VPCs per consentire una crescita pianificata dell'utilizzo.

Passaggi dell'implementazione

- Rileva CIDR il consumo di corrente (ad esempio, VPCs e le sottoreti).
  - Utilizza API le operazioni di servizio per raccogliere il consumo correnteCIDR.
  - Usa Amazon VPC IP Address Manager per scoprire le risorse.
- Misura l'utilizzo attuale delle sottoreti.
  - Utilizza API le operazioni di servizio per raccogliere sottoreti per ogni VPC regione.
  - Usa Amazon VPC IP Address Manager per scoprire le risorse.
- Registra l'uso attuale.
- Verifica se hai creato intervalli di indirizzi IP sovrapposti.
- Calcola la capacità inutilizzata.

- Individua gli intervalli di indirizzi IP sovrapposti. Puoi migrare verso una nuova gamma di indirizzi o prendere in considerazione l'utilizzo di tecniche come il NATgateway privato o AWS PrivateLinkse devi connettere gli intervalli sovrapposti.

Risorse

Best practice correlate:

- Protezione delle reti

Documenti correlati:

- APNPartner: partner che possono aiutarti a pianificare la tua rete
- Marketplace AWS per l'infrastruttura di rete
- Amazon Virtual Private Cloud Connectivity Options Whitepaper
- Multiple data center HA network connectivity
- Connecting Networks with Overlapping IP Ranges
- Che cos'è AmazonVPC?
- Che cos'èIPAM?

Video correlati:

- AWS re:Invent 2023 - VPC Design avanzati e nuove funzionalità
- AWS re:Invent 2019: architetture di riferimento per molti AWS Transit Gateway VPCs
- AWS re:Invent 2023 - Pronti per il futuro? Designing networks for growth and flexibility
- AWS re:Invent 2021 - {New Launch} Gestisci i tuoi indirizzi IP su larga scala AWS

Architettura del carico di lavoro

Questions

- REL3. Come si progetta l'architettura del servizio di carico di lavoro?
- REL4. Come si progettano le interazioni in un sistema distribuito per evitare errori?
- REL5. Come si progettano le interazioni in un sistema distribuito per mitigare o affrontare gli errori?

REL3. Come si progetta l'architettura del servizio di carico di lavoro?

Crea carichi di lavoro altamente scalabili e affidabili utilizzando un'architettura orientata ai servizi () o un'architettura di microservizi. SOA L'architettura orientata ai servizi (SOA) è la pratica di rendere i componenti software riutilizzabili tramite interfacce di servizio. L'architettura dei microservizi va oltre, per rendere i componenti più piccoli e semplici.

Best practice

- REL03-BP01 Scegli come segmentare il tuo carico di lavoro
- REL03-BP02 Crea servizi incentrati su domini e funzionalità aziendali specifici
- REL03-BP03 Fornire contratti di assistenza per API

REL03-BP01 Scegli come segmentare il tuo carico di lavoro

La segmentazione del carico di lavoro è importante nel determinare i requisiti di resilienza dell'applicazione. L'architettura monolitica va evitata, se possibile. Valuta invece con particolare attenzione quali componenti dell'applicazione possono essere suddivisi in microservizi. A seconda dei requisiti dell'applicazione, questa potrebbe finire per essere una combinazione di un'architettura orientata ai servizi () con microservizi, ove possibile. SOA I carichi di lavoro stateless sono più idonei all'implementazione come microservizi.

Risultato desiderato: i carichi di lavoro devono essere supportabili, scalabili e caratterizzati dal maggiore accoppiamento debole possibile.

Quando scegli come segmentare il carico di lavoro, trova il giusto compromesso tra i vantaggi e le complessità. Ciò che è giusto per un nuovo prodotto al primo lancio è diverso dai requisiti di un carico di lavoro creato per scalare le risorse. Durante la rifattorizzazione di un monolito esistente, dovrai considerare la capacità dell'applicazione di supportare la suddivisione in servizi stateless. La suddivisione dei servizi in elementi più piccoli consente a team ristretti e ben definiti di svilupparli e gestirli. Tuttavia, servizi di piccole dimensioni possono introdurre complessità, che includono un eventuale aumento della latenza, un debug più complesso e un maggiore carico operativo.

Anti-pattern comuni:

- Il microservizio Death Star rappresenta una situazione in cui i componenti atomici diventano così interdipendenti che un guasto verificatosi in un componente genera un guasto molto più grande, rendendo i componenti rigidi e fragili se considerati come monolito.

Vantaggi dell'adozione di questa best practice:

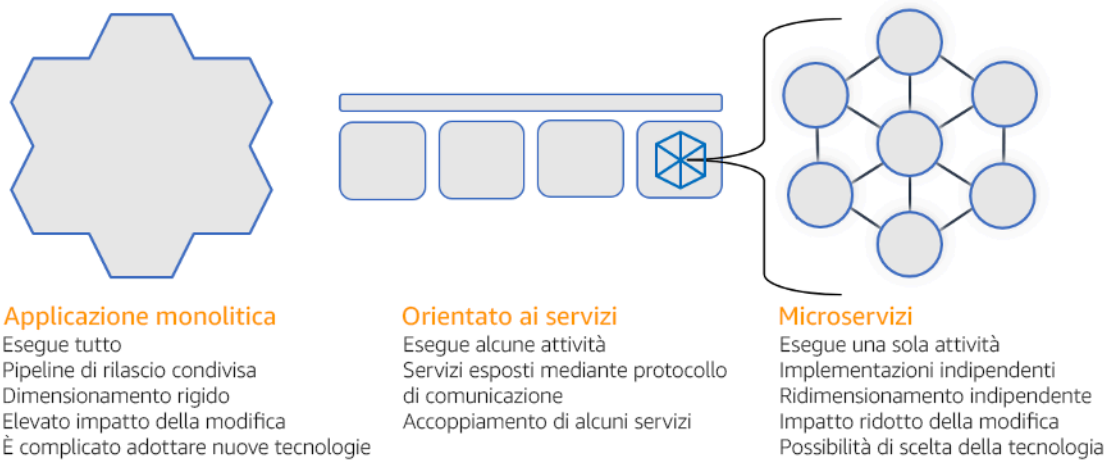
- Segmenti più specifici comportano maggiore agilità, flessibilità organizzativa e scalabilità.
- Riduzione dell'impatto derivante dall'interruzione dei servizi.
- I componenti dell'applicazione possono avere requisiti di disponibilità diversi, che a loro volta possono essere supportati da una segmentazione più atomica.
- Responsabilità ben definite per i team che supportano il carico di lavoro.

Livello di rischio associato se questa best practice non fosse adottata: elevato

Guida all'implementazione

Scegli il tipo di architettura in base al tipo di segmentazione del carico di lavoro. Scegliete un'SOAarchitettura a microservizi (o, in alcuni rari casi, un'architettura monolitica). Anche se scegli di iniziare con un'architettura monolitica, devi assicurarti che sia modulare e che alla fine possa evolversi verso i nostri microservizi man mano che il prodotto cresce con l'SOAadozione da parte degli utenti. SOAe i microservizi offrono rispettivamente una segmentazione più ridotta, che è preferibile in un'architettura moderna, scalabile e affidabile, ma ci sono dei compromessi da considerare, soprattutto quando si implementa un'architettura di microservizi.

Uno dei principali compromessi è che ora disponi di un'architettura di calcolo distribuita che può rendere più difficile il raggiungimento dei requisiti di latenza degli utenti ed è presente un'ulteriore complessità nel debug e nel tracciamento delle interazioni degli utenti. Puoi utilizzare AWS X-Ray per risolvere questo problema. Un altro effetto da considerare è l'aumento della complessità operativa man mano che aumenta il numero di applicazioni che gestisci, che richiede l'implementazione di più componenti di indipendenza.



Architettura monolitica, orientata ai servizi e di microservizi

Passaggi dell'implementazione

- Determina l'architettura più opportuna per rifattorizzare o creare l'applicazione. SOAe i microservizi offrono rispettivamente una segmentazione più piccola, preferibile come architettura moderna, scalabile e affidabile. SOApuò essere un buon compromesso per ottenere una segmentazione più piccola evitando al contempo alcune delle complessità dei microservizi. Per ulteriori dettagli, consulta Microservice Trade-Offs.
- Se il carico di lavoro è adatto e la tua organizzazione può supportarla, è consigliabile utilizzare un'architettura di microservizi per ottenere la massima agilità e affidabilità. Per ulteriori dettagli, consulta Implementazione dei microservizi su. AWS
- Valuta l'idea di attenerti al modello Strangler Fig per rifattorizzare un monolite in componenti più piccoli. Ciò comporta la sostituzione graduale di componenti applicativi specifici con nuove applicazioni e servizi. AWS Migration Hub Refactor Spaces funge da punto di partenza per procedere a rifattorizzare in modo incrementale. Per ulteriori dettagli, consulta Seamlessly migrate on-premises legacy workloads using a strangler pattern.
- L'implementazione dei microservizi può richiedere un meccanismo di rilevamento dei servizi per consentire a questi servizi distribuiti di comunicare tra loro. AWS App Meshpuò essere utilizzato con architetture orientate ai servizi per fornire l'individuazione e l'accesso affidabili ai servizi. AWS Cloud Mappuò essere utilizzato anche per l'individuazione dinamica dei servizi. DNS
- Se stai migrando da un monolite a, Amazon SOA MQ può aiutarti a colmare il divario come bus di servizio durante la riprogettazione delle applicazioni legacy nel cloud.

- Per i monoliti esistenti con un unico database condiviso, scegli come riorganizzare i dati in segmenti più piccoli. Questa riorganizzazione può avvenire per business unit, schema di accesso o struttura dei dati. A questo punto del processo di refactoring, dovresti scegliere di procedere con un tipo di database relazionale o non relazionale (No). SQL Per maggiori dettagli, consulta From to No. SQL SQL

Livello di impegno per il piano di implementazione: elevato

Risorse

Best practice correlate:

- REL03-BP02 Crea servizi incentrati su domini e funzionalità aziendali specifici

Documenti correlati:

- Amazon API Gateway: configurazione di un REST API utilizzo di Open API
- Cosa si intende per SOA (architettura orientata ai servizi)?
- Bounded Context (un modello centrale in Domain-Driven Design)
- Implementazione di microservizi su AWS
- Microservice Trade-Offs
- Microservices - a definition of this new architectural term
- Microservizi attivi AWS
- Che cos'è AWS App Mesh?

Esempi correlati:

- Iterative App Modernization Workshop

Video correlati:

- Offrire l'eccellenza con i microservizi attivi AWS

REL03-BP02 Crea servizi incentrati su domini e funzionalità aziendali specifici

Le architetture orientate ai servizi (SOA) definiscono servizi con funzioni ben delineate definite in base alle esigenze aziendali. I microservizi utilizzano modelli di dominio e contesto delimitato per tracciare i limiti dei servizi lungo i confini del contesto aziendale. Concentrarsi sui domini e sulle funzionalità aziendali aiuta i team a definire requisiti di affidabilità indipendenti per i propri servizi. I contesti delimitati isolano e incapsulano la logica aziendale, consentendo ai team di ragionare meglio su come gestire gli errori.

Risultato desiderato: ingegneri e parti interessate aziendali definiscono congiuntamente contesti delimitati e li utilizzano per progettare sistemi come servizi che soddisfano funzioni aziendali specifiche. Questi team utilizzano pratiche consolidate come l'event storming per definire i requisiti. Le nuove applicazioni sono concepite come servizi con confini ben definiti e con accoppiamento debole. I monoliti esistenti vengono scomposti in contesti limitati e i progetti di sistema si spostano verso architetture a microservizi. SOA In caso di rifattorizzazione dei monoliti, vengono applicati approcci consolidati come contesti a bolle e schemi di decomposizione dei monoliti.

I servizi orientati al dominio vengono eseguiti come uno o più processi che non condividono lo stato. Rispondono in modo indipendente alle fluttuazioni della domanda e gestiscono gli scenari di errore alla luce dei requisiti specifici del dominio.

Anti-pattern comuni:

- I team sono formati su domini tecnici specifici come UI e UX, middleware (software intermediario) o database anziché su domini aziendali specifici.
- Le applicazioni coprono le responsabilità di dominio. I servizi che coprono contesti delimitati possono essere più difficili da gestire, richiedere maggiori sforzi di test ed esigere la partecipazione di più team di dominio agli aggiornamenti software.
- Le dipendenze a livello di dominio, come le librerie di entità di dominio, sono condivise tra i servizi, in modo che le modifiche per il dominio di un servizio richiedano modifiche ad altri domini dei servizi.
- I contratti di servizio e la logica aziendale non esprimono le entità in un linguaggio di dominio comune e coerente, con il risultato di livelli di traduzione che complicano i sistemi e aumentano le attività di debug.

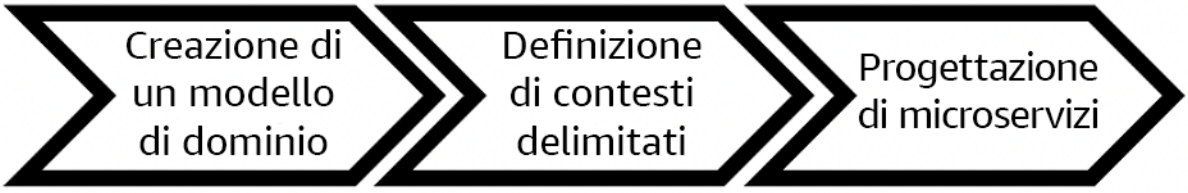
Vantaggi dell'adozione di questa best practice: le applicazioni sono progettate come servizi indipendenti limitati da domini aziendali e utilizzano un linguaggio aziendale comune. I servizi sono

testabili e implementabili in modo indipendente. I servizi soddisfano i requisiti di resilienza specifici del dominio implementato.

Livello di rischio associato se questa best practice non fosse adottata: elevato

Guida all'implementazione

La progettazione basata sul dominio (DDD) è l'approccio fondamentale per la progettazione e la creazione di software in base ai domini aziendali. È utile utilizzare un framework esistente quando si creano servizi incentrati sui domini aziendali. Quando si utilizzano applicazioni monolitiche esistenti, è possibile sfruttare i modelli di decomposizione che forniscono tecniche consolidate per modernizzare le applicazioni in servizi.



Progettazione basata sul dominio

Passaggi dell'implementazione

- I team possono organizzare workshop di event storming per identificare rapidamente eventi, comandi, aggregati e domini in un formato leggero simile a quello delle note adesive.
- Una volta create le entità e le funzioni di dominio in un contesto di dominio, puoi suddividere il dominio in servizi mediante il contesto delimitato, che raggruppa entità con funzionalità e attributi simili. Con il modello diviso in contesti, emerge un modello su come delimitare i microservizi.
  - Ad esempio, le entità del sito Web Amazon.com possono includere elementi quali pacchetti, distribuzione, pianificazione, prezzo, sconto e valuta.
  - Il pacchetto, la distribuzione e la pianificazione sono raggruppati nel contesto di spedizione, mentre il prezzo, lo sconto e la valuta sono raggruppati nel contesto dei prezzi.
- La scomposizione dei monoliti in microservizi delinea i modelli per rifattorizzare i microservizi. L'utilizzo di modelli per la decomposizione in base a capacità aziendale, sottodominio o transazione si allinea bene agli approcci basati sul dominio.
- Le tecniche tattiche, come il bubble context, consentono di introdurre applicazioni esistenti o legacy senza riscritture iniziali e impegni completi DDD in tal senso. DDD In un approccio basato sul contesto delle bolle, si crea un contesto ristretto e delimitato mediante un livello di mappatura e

coordinamento dei servizi o il livello anticorruzione, che protegge il modello di dominio appena definito dalle influenze esterne.

Dopo aver eseguito l'analisi del dominio e definito le entità e i contratti di servizio, i team possono sfruttare i AWS servizi per implementare la loro progettazione basata sul dominio come servizi basati sul cloud.

- Inizia a sviluppare definendo test che applichino le regole aziendali del tuo dominio. Lo sviluppo basato sui test (TDD) e lo sviluppo basato sul comportamento (BDD) aiutano i team a mantenere i servizi concentrati sulla risoluzione dei problemi aziendali.
- Seleziona i servizi AWS ideali per i requisiti del tuo dominio aziendale e l'architettura dei microservizi:
  - AWS Serverless consente al team di concentrarsi su una logica di dominio specifica anziché sulla gestione di server e infrastrutture.
  - I container in AWS semplificano la gestione della tua infrastruttura, in modo da poterti concentrare sui requisiti del tuo dominio.
  - I database dedicati ti aiutano ad adattare i requisiti del tuo dominio al tipo di database più idoneo.
- La creazione di architetture esagonali in AWS delinea un framework per integrare la logica aziendale nei servizi che funzionano a ritroso da un dominio aziendale per soddisfare i requisiti funzionali e, quindi, per collegare adattatori di integrazione. I modelli che separano i dettagli dell'interfaccia dalla logica aziendale con AWS i servizi aiutano i team a concentrarsi sulle funzionalità del dominio e a migliorare la qualità del software.

Risorse

Best practice correlate:

- REL03-BP01 Scegli come segmentare il tuo carico di lavoro
- REL03-BP03 Fornire contratti di assistenza per API

Documenti correlati:

- AWS Microservizi
- Implementazione di microservizi su AWS
- How to break a Monolith into Microservices

- Come iniziare DDD quando sei circondato da sistemi legacy
- Domain-Driven Design: Tackling Complexity in the Heart of Software
- Costruire architetture esagonali su AWS
- Decomposing monoliths into microservices
- Event Storming
- Messages Between Bounded Contexts
- Microservices
- Sviluppo basato su test
- Sviluppo basato sul comportamento

Esempi correlati:

- Progettazione di microservizi nativi per il cloud su (da/) AWS DDD EventStormingWorkshop

Strumenti correlati:

- Cloud AWS Database
- Serverless attivo AWS
- Contenitori presso AWS

REL03-BP03 Fornire contratti di assistenza per API

I contratti di assistenza sono accordi documentati tra API produttori e consumatori definiti in una definizione leggibile da una macchinaAPI. Una strategia di controllo delle versioni contrattuali consente ai consumatori di continuare a utilizzare le applicazioni esistenti API e di migrare le proprie applicazioni a una versione più recente quando sono pronte. API L'implementazione da parte del produttore può avvenire in qualsiasi momento, purché il processo sia conforme al contratto. I team di assistenza possono utilizzare lo stack tecnologico di loro scelta per soddisfare il contratto. API

Risultato desiderato: le applicazioni create con architetture orientate ai servizi o ai microservizi sono in grado di funzionare in modo indipendente pur avendo una dipendenza di runtime integrata. Le modifiche apportate a un API consumatore o a un produttore non interrompono la stabilità dell'intero sistema quando entrambe le parti seguono un contratto comune. API I componenti che comunicano tramite servizio APIs possono eseguire rilasci funzionali indipendenti, aggiornamenti alle dipendenze di runtime o eseguire il failover su un sito di disaster recovery (DR) con un impatto reciproco minimo

o nullo. Inoltre, i servizi discreti sono in grado di scalare in modo indipendente assorbendo la richiesta di risorse senza che gli altri servizi debbano ridurre orizzontalmente di conseguenza.

Anti-pattern comuni:

- Creazione di servizi APIs senza schemi fortemente tipizzati. Ciò comporta APIs che non può essere utilizzato per generare API associazioni e payload che non possono essere convalidati programmaticamente.
- Non adottare una strategia di controllo delle versioni, che costringerebbe gli API utenti ad aggiornare e rilasciare o fallire quando i contratti di assistenza si evolvono.
- Messaggi di errore che divulgano dettagli sull'implementazione del servizio sottostante anziché descrivere errori di integrazione nel contesto e nel linguaggio del dominio.
- Non utilizzare API contratti per sviluppare casi di test e API implementazioni fittizie per consentire test indipendenti dei componenti del servizio.

Vantaggi derivanti dall'adozione di questa best practice: i sistemi distribuiti composti da componenti che comunicano tramite contratti di API assistenza possono migliorare l'affidabilità. Gli sviluppatori possono individuare potenziali problemi nelle prime fasi del processo di sviluppo grazie al controllo del tipo durante la compilazione per verificare che le richieste e le risposte siano conformi al API contratto e che i campi obbligatori siano presenti. APIi contratti forniscono una chiara interfaccia di autodocumentazione APIs e forniscono una migliore interoperabilità tra diversi sistemi e linguaggi di programmazione.

Livello di rischio associato se questa best practice non fosse adottata: medio

Guida all'implementazione

Dopo aver identificato i domini aziendali e determinato la segmentazione del carico di lavoro, puoi sviluppare il tuo servizio. APIs Innanzitutto, definisci contratti di assistenza leggibili automaticamente e poi implementa una strategia di controllo delle APIs versioni. API Quando sei pronto per integrare i servizi tramite protocolli comuni come REST GraphQL o eventi asincroni, puoi incorporare AWS servizi nella tua architettura per integrare i componenti con contratti fortemente tipizzati. API

AWS APIservizi per contratti di assistenza

Incorpora AWS servizi tra cui Amazon API Gateway e Amazon EventBridge nella tua architettura per utilizzare i contratti di API servizio nella tua applicazione. AWS AppSync Amazon API Gateway ti aiuta a integrarti con AWS servizi nativi diretti e altri servizi Web. APIGateway supporta le APIspecifiche e il controllo delle versioni Open. AWS AppSync è un endpoint GraphQL gestito

che puoi configurare definendo uno schema GraphQL per definire un'interfaccia di servizio per query, mutazioni e sottoscrizioni. Amazon EventBridge utilizza schemi di eventi per definire eventi e generare associazioni di codice per i tuoi eventi.

Passaggi dell'implementazione

- Innanzitutto, definisci un contratto per il tuo. API Un contratto esprimerà le capacità di un e API definirà oggetti di dati e campi fortemente tipizzati per l'APIinput e l'output.
- Quando esegui la configurazione APIs in API Gateway, puoi importare ed esportare API le specifiche aperte per i tuoi endpoint.
  - L'importazione di una API definizione aperta semplifica la creazione della propria definizione API e può essere integrata con l' AWS infrastruttura come strumenti di codice come la e. AWS Serverless Application ModelAWS Cloud Development Kit (AWS CDK)
  - L'esportazione di una API definizione semplifica l'integrazione con gli strumenti di API test e fornisce ai consumatori di servizi una specifica di integrazione.
- Puoi definire e gestire GraphQL APIs AWS AppSync definendo un file di schema GraphQL per generare l'interfaccia del contratto e semplificare l'interazione con REST modelli complessi, più tabelle di database o servizi legacy.
- AWS Amplifyi progetti integrati AWS AppSync generano file di JavaScript query fortemente tipizzati da utilizzare nella tua applicazione e una libreria client AWS AppSync GraphQL per le tabelle Amazon DynamoDB.
- Quando utilizzi gli eventi di servizio di Amazon EventBridge, gli eventi aderiscono a schemi già esistenti nel registro degli schemi o definiti con Open API Spec. Con uno schema definito nel registro, puoi anche generare associazioni client dal contratto dello schema per integrare il codice con gli eventi.
- Estendere o modificare il tuo. API L'estensione di un API è un'opzione più semplice quando si aggiungono campi che possono essere configurati con campi opzionali o valori predefiniti per i campi obbligatori.
  - JSONi contratti basati su protocolli come REST GraphQL possono essere adatti per l'estensione del contratto.
  - XMLi contratti basati su protocolli, ad esempio, SOAP dovrebbero essere testati con i consumatori di servizi per determinare la fattibilità dell'estensione del contratto.
- Quando si effettua il versionamento di unAPI, è consigliabile implementare il controllo delle versioni proxy, in cui viene utilizzata una facciata per supportare le versioni in modo che la logica possa essere mantenuta in un'unica base di codice.

- Con API Gateway puoi utilizzare le mappature di richiesta e risposta per semplificare l'assorbimento delle modifiche contrattuali, stabilendo una facciata per fornire valori predefiniti per nuovi campi o per eliminare i campi rimossi da una richiesta o risposta. Con questo approccio, il servizio sottostante può avere un'unica base di codice.

Risorse

Best practice correlate:

- REL03-BP01 Scegli come segmentare il tuo carico di lavoro
- REL03-BP02 Crea servizi incentrati su domini e funzionalità aziendali specifici
- REL04-BP02 Implementare dipendenze liberamente accoppiate
- REL05-BP03 Controlla e limita le chiamate di nuovo tentativo
- REL05-BP05 Imposta i timeout dei client

Documenti correlati:

- Che cos'è una API (interfaccia di programmazione delle applicazioni)?
- Implementazione di microservizi su AWS
- Microservice Trade-Offs
- Microservices - a definition of this new architectural term
- Microservizi attivi AWS
- Utilizzo delle estensioni API Gateway to Open API
- Apri API - Specificazione
- GraphQL: schemi e tipi
- Associazioni di EventBridge codice Amazon

Esempi correlati:

- Amazon API Gateway: configurazione di un REST API utilizzo di Open API
- Da Amazon API Gateway all'applicazione Amazon CRUD DynamoDB tramite Open API
- Modelli di integrazione delle applicazioni moderni nell'era senza server: API Gateway Service Integration

- Implementazione del controllo delle versioni API Gateway basato su header con Amazon CloudFront
- AWS AppSync: Building a client application

Video correlati:

- Utilizzo di Open API in AWS SAM per gestire Gateway API

Strumenti correlati:

- Amazon API Gateway
- AWS AppSync
- Amazon EventBridge

REL4. Come si progettano le interazioni in un sistema distribuito per evitare errori?

I sistemi distribuiti si basano sulle reti di comunicazione per interconnettere i componenti, ad esempio server o servizi. Il carico di lavoro deve funzionare in modo affidabile nonostante la perdita o la latenza dei dati su queste reti. I componenti del sistema distribuito devono funzionare in modo da non influire negativamente su altri componenti o sul carico di lavoro. Queste best practice prevengono gli errori e migliorano il tempo medio tra i guasti (). MTBF

Best practice

- REL04-BP01 Identifica il tipo di sistemi distribuiti da cui dipendi
- REL04-BP02 Implementare dipendenze liberamente accoppiate
- REL04-BP03 Fai un lavoro costante
- REL04-BP04 Rendi tutte le risposte idempotenti

REL04-BP01 Identifica il tipo di sistemi distribuiti da cui dipendi

I sistemi distribuiti possono essere sincroni, asincroni o batch. I sistemi sincroni devono elaborare le richieste il più rapidamente possibile e comunicare tra loro effettuando chiamate di richiesta e risposta sincrone utilizzando i protocolli HTTP /S o remote procedure call (). REST RPC I sistemi asincroni comunicano tra loro scambiando i dati in modo asincrono tramite un servizio intermediario senza associare singoli sistemi. I sistemi batch ricevono un grande volume di dati di input, eseguono i processi di dati automatizzati senza intervento umano e generano i dati di output.

Risultato desiderato: progettazione di un carico di lavoro in grado di interagire in modo efficace con dipendenze sincrone, asincrone e batch.

Anti-pattern comuni:

- Il carico di lavoro attende a tempo indeterminato una risposta dalle dipendenze, con eventuale timeout del client del carico di lavoro, senza informazioni sulla ricezione della richiesta.
- Il carico di lavoro utilizza una catena di sistemi dipendenti che effettuano chiamate reciproche in modo sincrono. A tal fine, ogni sistema deve essere disponibile ed elaborare correttamente la richiesta prima che l'intera catena possa essere completata, con conseguenti comportamenti e disponibilità complessiva potenzialmente fragili.
- Il carico di lavoro comunica con le dipendenze in modo asincrono e si basa sul concetto di distribuzione garantita dei messaggi esattamente una volta, quando spesso è ancora possibile ricevere messaggi duplicati.
- Il carico di lavoro non utilizza strumenti di pianificazione batch adeguati e consente l'esecuzione simultanea dello stesso processo batch.

Vantaggi dell'adozione di questa best practice: non è insolito che un determinato carico di lavoro implementi uno o più stili di comunicazione tra sincroni, asincroni e batch. Questa best practice consente di identificare i diversi compromessi associati a ogni stile di comunicazione per rendere il carico di lavoro in grado di tollerare interruzioni in tutte le sue dipendenze.

Livello di rischio associato se questa best practice non fosse adottata: elevato

Guida all'implementazione

Le sezioni seguenti contengono le linee guida per l'implementazione generali e specifiche di ogni tipo di dipendenza.

Informazioni generali

- Assicurati che gli obiettivi a livello di servizio (SLOs) in termini di prestazioni e affidabilità offerti dalle tue dipendenze soddisfino i requisiti di prestazioni e affidabilità del tuo carico di lavoro.
- Utilizza i servizi di osservabilitàAWS per monitorare i tempi di risposta e i tassi di errore così da verificare che la tua dipendenza fornisca un servizio ai livelli richiesti dal carico di lavoro.
- Individua le potenziali sfide che il carico di lavoro può affrontare quando comunica con le dipendenze. I sistemi distribuiti presentano un'ampia gamma di sfide in grado di far aumentare

complessità dell'architettura, carico operativo e costi. Le sfide più comuni includono latenza, interruzioni della rete, perdita dei dati, scalabilità e ritardo nella replica dei dati.

- Implementa una gestione e una creazione di log affidabili degli errori per risolvere i problemi quando si verificano quelli legati alle dipendenze.

Dipendenza sincrona

Nelle comunicazioni sincrone, il carico di lavoro invia una richiesta alla dipendenza e blocca l'operazione in attesa della risposta. Quando la dipendenza riceve la richiesta, cerca di gestirla il prima possibile e invia una risposta al carico di lavoro. Una sfida significativa con la comunicazione sincrona è rappresentata dall'accoppiamento temporale, che richiede che il carico di lavoro e le sue dipendenze siano disponibili nello stesso momento. Quando il carico di lavoro deve comunicare in modo sincrono con le dipendenze, valuta le seguenti linee guida:

- Il carico di lavoro non deve fare affidamento su più dipendenze sincrone per eseguire una singola funzione. Questa catena di dipendenze aumenta la fragilità complessiva perché tutte le dipendenze nel percorso devono essere disponibili affinché la richiesta venga completata correttamente.
- Quando una dipendenza non è integra o non è disponibile, applica le strategie di gestione degli errori e riprova. Evita di usare un comportamento bimodale. Il comportamento bimodale si verifica quando il carico di lavoro presenta un comportamento diverso in modalità normale e in modalità di guasto. Per maggiori dettagli sul comportamento bimodale, consulta REL11-BP05 Utilizzare la stabilità statica per prevenire il comportamento bimodale.
- Tieni presente che anticipare l'errore (fail fast) è meglio che far aspettare il carico di lavoro. Ad esempio, la guida per gli sviluppatori AWS Lambda illustra come gestire tentativi ed errori nel richiamare le funzioni Lambda.
- Imposta i timeout per le chiamate delle dipendenze da parte del carico di lavoro. Questa tecnica evita di aspettare troppo a lungo o all'infinito una risposta. Per una discussione utile su questo problema, consulta Tuning AWS Java SDK HTTP request settings for latency-aware applications Amazon DynamoDB.
- Riduci al minimo il numero di chiamate effettuate dal carico di lavoro alla dipendenza per soddisfare una singola richiesta. Le lunghe chiamate aumentano l'associazione e la latenza.

Dipendenza sincrona

Per disaccoppiare temporaneamente il carico di lavoro dalla dipendenza, è necessario che comunichino in modo asincrono. Con l'approccio asincrono, il carico di lavoro può continuare

qualsiasi altra elaborazione senza dover attendere che la dipendenza o la catena di dipendenze invii la risposta.

Quando il carico di lavoro deve comunicare in modo asincrono con la dipendenza, tieni conto delle seguenti indicazioni:

- Determina in base al caso d'uso e ai requisiti se utilizzare la messaggistica o lo streaming di eventi. La messaggistica consente al carico di lavoro di comunicare con la relativa dipendenza, inviando e ricevendo messaggi tramite un broker di messaggi. Lo streaming di eventi consente al carico di lavoro e alla relativa dipendenza di utilizzare un servizio di streaming per la pubblicazione e l'abbonamento a eventi, forniti come flussi continui di dati, da elaborare il prima possibile.

- La messaggistica e lo streaming di eventi gestiscono i messaggi in modo diverso, quindi devi stabilire i compromessi in base a:
  - Priorità dei messaggi: i broker di messaggi sono in grado di elaborare messaggi ad alta priorità prima di quelli normali. Nello streaming di eventi, tutti i messaggi presentano la stessa priorità.
  - Consumo di messaggi: i broker di messaggi garantiscono la ricezione del messaggio da parte dei consumatori. Gli utenti che utilizzano lo streaming di eventi devono tenere traccia dell'ultimo messaggio letto.
  - Ordinamento dei messaggi: con la messaggistica, la ricezione dei messaggi nell'ordine esatto in cui vengono inviati non è garantita a meno che non si utilizzi un approccio (). first-in-first-out FIFO Lo streaming di eventi mantiene sempre l'ordine in cui i dati sono stati prodotti.
  - Eliminazione dei messaggi: con la messaggistica, il consumatore deve eliminare il messaggio dopo la relativa elaborazione. Il servizio di streaming di eventi aggiunge il messaggio a un flusso e lo conserva fino alla scadenza del periodo di conservazione del messaggio. Questa policy di eliminazione rende lo streaming di eventi adatto alla riproduzione dei messaggi.
- Definisci in che modo il carico di lavoro riconosce il completamento del lavoro della dipendenza. Ad esempio, quando il carico di lavoro procede a richiamare una funzione Lambda in modo asincrono, Lambda inserisce la richiesta in una coda e restituisce una risposta di esito positivo senza ulteriori informazioni. Al termine dell'elaborazione, la funzione Lambda può inviare il risultato a una destinazione, configurabile in base all'esito positivo o negativo.
- Crea il tuo carico di lavoro per gestire i messaggi duplicati utilizzando l'idempotenza. Con l'idempotenza i risultati del carico di lavoro non cambiano anche se il carico di lavoro viene generato più volte per lo stesso messaggio. È importante sottolineare che i servizi di messaggistica o streaming recapiteranno di nuovo un messaggio in caso di errore di rete o mancata ricezione della conferma.

- Se il carico di lavoro non riceve una risposta dalla dipendenza, deve inviare nuovamente la richiesta. Valuta la possibilità di limitare il numero di tentativi per preservare il carico di lavoroCPU, la memoria e le risorse di rete per gestire altre richieste. La documentazione AWS Lambda illustra come gestire gli errori di invocazione asincrona.
- Utilizza gli strumenti di osservabilità, debug e monitoraggio adeguati per gestire e usare la comunicazione asincrona del carico di lavoro con le relative dipendenze. Puoi usare Amazon CloudWatch per monitorare i servizi di messaggistica e streaming di eventi. Puoi anche dotare di strumenti il tuo carico di lavoro con AWS X-Ray per ottenere informazioni utili rapidamente per la risoluzione dei problemi.

Dipendenza dal batch

I sistemi batch acquisiscono i dati di input, avviano una serie di processi per elaborarli e producono i dati di output, senza intervento manuale. A seconda delle dimensioni dei dati, i processi possono durare da minuti a diversi giorni in alcuni casi. Quando il carico di lavoro comunica con la dipendenza batch, tieni conto delle seguenti indicazioni:

- Definisci la finestra temporale in cui il carico di lavoro deve eseguire il processo batch. Puoi impostare un modello di ricorrenza per il carico di lavoro per richiamare il sistema batch, ad esempio ogni ora o alla fine di ogni mese.
- Determina la posizione dei dati di input e di output elaborati. Scegli un servizio di storage, come Amazon Simple Storage Services (Amazon S3), Amazon Elastic File System (AmazonEFS) e Amazon FSx for Lustre, che consenta al tuo carico di lavoro di leggere e scrivere file su larga scala.
- Se il tuo carico di lavoro deve richiamare più processi batch, puoi AWS Step Functionssfruttarlo per semplificare l'orchestrazione dei processi batch eseguiti in locale o in locale. AWS Questo progetto di esempio mostra l'orchestrazione di processi batch utilizzando Step Functions, AWS Batch e Lambda.
- Monitora i processi batch per individuare eventuali anomalie, ad esempio un processo che richiede più tempo del dovuto per essere completato. Puoi usare strumenti come CloudWatchContainer Insights per monitorare ambienti e lavori. AWS Batch In tal caso, il carico di lavoro interrompe l'inizio del processo successivo e comunica l'eccezione al team competente.

Risorse

Documenti correlati:

- Cloud AWS Operazioni: monitoraggio e osservabilità

- Amazon Builders' Library: difficoltà dei sistemi distribuiti
- REL11-BP05 Usa la stabilità statica per prevenire comportamenti bimodali
- AWS Lambda Guida per gli sviluppatori: gestione degli errori e tentativi automatici in AWS Lambda
- Ottimizzazione delle impostazioni delle SDK HTTP richieste AWS Java per le applicazioni Amazon DynamoDB sensibili alla latenza
- Messaggi AWS
- Cosa sono i flussi di dati?
- AWS Lambda Guida per gli sviluppatori: invocazione asincrona
- Amazon Simple Queue ServiceFAQ: FIFO code
- Amazon Kinesis Data Streams Developer Guide: Handling Duplicate Records
- Guida per gli sviluppatori di Amazon Simple Queue Service: CloudWatch metriche disponibili per Amazon SQS
- Guida per sviluppatori di Amazon Kinesis Data Streams: monitoraggio del servizio Amazon Kinesis Data Streams con Amazon CloudWatch
- AWS X-Ray Guida per sviluppatori: concetti AWS X-Ray
- AWS Esempi sull'app GitHub AWS Step functions Complex Orchestrator
- AWS Batch Guida per l'utente: Container Insights AWS Batch CloudWatch

Video correlati:

- AWS Summit SF 2022 - Osservabilità completa e monitoraggio delle applicazioni con (0) AWS COP31

Strumenti correlati:

- Amazon CloudWatch
- CloudWatch Registri Amazon
- AWS X-Ray
- Amazon Simple Storage Service (Amazon S3)
- Amazon Elastic File System (AmazonEFS)
- Amazon FSx per Lustre
- AWS Step Functions

- AWS Batch

REL04-BP02 Implementare dipendenze liberamente accoppiate

Le dipendenze come sistemi di accodamento, sistemi di streaming, flussi di lavoro e bilanciatori del carico sono con accoppiamento debole. L'accoppiamento debole aiuta a isolare il comportamento di un componente dagli altri componenti che dipendono da esso, aumentando la resilienza e l'agilità.

Il disaccoppiamento delle dipendenze, come i sistemi di coda, quelli di streaming e i flussi di lavoro, favorisce la riduzione al minimo dell'impatto sul sistema di modifiche o guasti. Tale separazione isola il comportamento di un componente dall'impatto sugli altri dipendenti dallo stesso, migliorando resilienza e agilità.

Nei sistemi con accoppiamento stretto, le modifiche a un componente possono richiedere modifiche agli altri componenti basati su di esso, con conseguente riduzione delle prestazioni di tutti i componenti. L'accoppiamento debole interrompe questa dipendenza, in modo che i componenti dipendenti debbano conoscere solo l'interfaccia con versione e pubblicata. L'implementazione di un accoppiamento debole tra dipendenze isola un errore all'interno di una dipendenza affinché non influenzi l'altra.

L'accoppiamento debole consente di modificare il codice o aggiungere funzionalità a un componente riducendo al minimo il rischio per gli altri componenti che dipendono da esso. Garantisce inoltre una resilienza granulare a livello di componente in cui è possibile aumentare orizzontalmente o persino modificare l'implementazione sottostante della dipendenza.

Per migliorare ulteriormente la resilienza tramite accoppiamento debole, rendi le interazioni dei componenti asincrone laddove possibile. Questo modello è idoneo a qualsiasi interazione che non richieda una risposta immediata e laddove la conferma della registrazione di una richiesta sia sufficiente. Include un componente che genera eventi e un altro che li utilizza. I due componenti non si integrano tramite un' point-to-pointinterazione diretta, ma di solito attraverso un livello di storage intermedio durevole, come una SQS coda Amazon, una piattaforma di dati di streaming come Amazon Kinesis o. AWS Step Functions

Figura 4: dipendenze come sistemi di accodamento e bilanciatori del carico con accoppiamento debole

Amazon mette in SQS coda e AWS Step Functions sono solo due modi per aggiungere uno strato intermedio per l'accoppiamento libero. Le architetture basate sugli eventi possono anche essere

create utilizzando Cloud AWS Amazon EventBridge, che può astrarre i clienti (produttori di eventi) dai servizi su cui fanno affidamento (consumatori di eventi). Amazon Simple Notification Service (AmazonSNS) è una soluzione efficace quando è necessaria una messaggistica basata su push ad alta velocità. many-to-many Utilizzando SNS gli argomenti di Amazon, i tuoi sistemi di pubblicazione possono inviare messaggi a un gran numero di endpoint di abbonati per l'elaborazione parallela.

Mentre le code offrono diversi vantaggi, nella maggior parte dei sistemi hard real-time, le richieste più vecchie di una soglia temporale (spesso secondi) dovrebbero essere considerate obsolete (il client ha abbandonato e non è più in attesa di una risposta) e non elaborate. In questo modo, è possibile elaborare invece le richieste più recenti (e probabilmente ancora valide).

Risultato desiderato: riduzione al minimo l'area della superficie in caso di guasto a livello di componente, supportando così diagnostica e risoluzione dei problemi, grazie all'implementazione di dipendenze con accoppiamento debole. Inoltre, semplifica i cicli di sviluppo, consentendo ai team di implementare le modifiche a livello modulare senza pregiudicare le prestazioni di altri componenti che dipendono da esso. Questo approccio offre la possibilità di aumentare orizzontalmente a livello di componente in base al fabbisogno di risorse, nonché di utilizzare un componente che contribuisce alla competitività in termini di costi.

Anti-pattern comuni:

- Implementazione di un carico di lavoro monolitico.
- Richiamo diretto APIs tra livelli di carico di lavoro senza possibilità di failover o elaborazione asincrona della richiesta.
- Accoppiamento stretto utilizzando dati condivisi. I sistemi con accoppiamento debole dovrebbero evitare di condividere i dati tramite database condivisi o altre forme di archiviazione di dati con accoppiamento stretto, che possono reintrodurre l'accoppiamento stretto e compromettere la scalabilità.
- Ignorare la contropressione. Il carico di lavoro dovrebbe essere in grado di rallentare o arrestare i dati in arrivo quando un componente non è in grado di elaborarli alla stessa velocità.

Vantaggi dell'adozione di questa best practice: l'accoppiamento debole aiuta a isolare il comportamento di un componente dagli altri componenti che dipendono da esso, aumentando la resilienza e l'agilità. L'errore in un componente è isolato dagli altri.

Livello di rischio associato se questa best practice non fosse adottata: elevato

Guida all'implementazione

Implementazione di dipendenze con accoppiamento debole Esistono varie soluzioni che consentono di creare applicazioni con accoppiamento debole. Questi includono servizi per l'implementazione di code completamente gestite, flussi di lavoro automatizzati, la reazione agli eventi e, APIs tra gli altri, che possono aiutare a isolare il comportamento dei componenti dagli altri componenti e, di conseguenza, aumentare la resilienza e l'agilità.

- Crea architetture basate sugli eventi: EventBridgeAmazon ti aiuta a creare architetture basate sugli eventi liberamente accoppiate e distribuite.
- Implementazione di code in sistemi distribuiti: puoi utilizzare Amazon Simple Queue Service SQS (Amazon) per integrare e disaccoppiare sistemi distribuiti.
- Containerizza i componenti come microservizi: i microservizi consentono ai team di creare applicazioni composte da piccoli componenti indipendenti che comunicano in modo ben definito. APIs Amazon Elastic Container Service (AmazonECS) e Amazon Elastic Kubernetes Service (EKSAmazon) possono aiutarti a iniziare a usare i container più velocemente.
- Gestisci i flussi di lavoro con Step Functions: Step Functions ti aiuta a coordinare più AWS servizi in flussi di lavoro flessibili.
- Sfrutta le architetture di messaggistica publish-subscribe (pub/sub): Amazon Simple Notification Service (Amazon SNS) fornisce il recapito dei messaggi dagli editori agli abbonati (noti anche come produttori e consumatori).

Passaggi dell'implementazione

- I componenti in un'architettura basata su eventi vengono avviati dagli eventi. Gli eventi sono azioni che si verificano in un sistema, ad esempio un utente che aggiunge un articolo a un carrello. Quando un'azione ha successo, viene generato un evento che attiva il successivo componente del sistema.
  - Creazione di applicazioni basate sugli eventi con Amazon EventBridge
  - AWS re:Invent 2022 - Progettazione di integrazioni basate sugli eventi con Amazon EventBridge
- I sistemi di messaggistica distribuiti sono composti da tre parti principali che devono essere implementate per un'architettura basata su code. Includono componenti del sistema distribuito, la coda utilizzata per il disaccoppiamento (distribuita sui SQS server Amazon) e i messaggi in coda. Un sistema tipico prevede produttori che inviano il messaggio alla coda e il consumatore che riceve il messaggio dalla coda. La coda archivia i messaggi su più SQS server Amazon per motivi di ridondanza.

  - SQSArchitettura Amazon di base
  - Send Messages Between Distributed Applications with Amazon Simple Queue Service
- I microservizi, se ben utilizzati, migliorano la manutenibilità e aumentano la scalabilità, poiché i componenti ad accoppiamento debole sono gestiti da team indipendenti. Consentono inoltre l'isolamento dei comportamenti in un unico componente in caso di modifiche.
  - Implementazione di microservizi su AWS
  - Let's Architect! Architecting microservices with containers
- Con AWS Step Functions puoi creare applicazioni distribuite, automatizzare i processi, orchestrare microservizi, tra le altre cose. L'orchestrazione di più componenti in un flusso di lavoro automatizzato consente di disaccoppiare le dipendenze nell'applicazione.
  - Crea un flusso di lavoro serverless con e AWS Step FunctionsAWS Lambda
  - Guida introduttiva con AWS Step Functions

Risorse

Documenti correlati:

- AmazonEC2: garantire l'idempotenza
- The Amazon Builders' Library: difficoltà dei sistemi distribuiti
- The Amazon Builders' Library: Reliability, constant work, and a good cup of coffee
- Che cos'è Amazon EventBridge?
- What Is Amazon Simple Queue Service?
- Break up with your monolith
- Orchestra i microservizi basati sulle code con Amazon AWS Step Functions SQS
- SQSArchitettura Amazon di base
- Queue-Based Architecture

Video correlati:

- AWS New York Summit 2019: introduzione alle architetture basate sugli eventi e ad Amazon (05) EventBridge MAD2
- AWS re:Invent 2018: Close Loops and Opening Minds: Come prendere il controllo di sistemi, grandi e piccoli ARC337 (include accoppiamento libero, lavoro costante, stabilità statica)

- AWS re:Invent 2019: Passaggio ad architetture basate sugli eventi (08) SVS3
- AWS re:Invent 2019: applicazioni scalabili senza server basate su eventi con Amazon e Lambda SQS
- AWS re:Invent 2022 - Progettazione di integrazioni basate sugli eventi con Amazon EventBridge
- AWS re:Invent 2017: Approfondimento e best practice su Elastic Load Balancing

REL04-BP03 Fai un lavoro costante

I sistemi possono presentare guasti quando si verificano modifiche rapide e di grandi dimensioni nel carico. Ad esempio, se il carico di lavoro effettua un controllo dell'integrità di migliaia di server deve inviare ogni volta lo stesso payload delle dimensioni (uno snapshot completo dello stato corrente). Indipendentemente dal fatto che non ci siano server guasti, o che lo siano tutti, il sistema di controllo dell'integrità esegue un lavoro costante con modifiche rapide e di piccole dimensioni.

Ad esempio, se il sistema di controllo dell'integrità monitora 100.000 server, il carico su di esso è nominale al di sotto del tasso di errore normalmente basso del server. Tuttavia, se un evento importante rendesse la metà di questi server non integra, il sistema di controllo dell'integrità sarebbe sovraccarico nel tentativo di aggiornare i sistemi di notifica e comunicare lo stato con i client. Pertanto, il sistema di controllo dell'integrità dovrebbe inviare ogni volta lo snapshot completo dello stato attuale. 100.000 stati di integrità del server, ciascuno rappresentato da un bit, equivarrebbero a un payload di soli 12,5 KB. Indipendentemente dal fatto che non ci siano server guasti, o che lo siano tutti, il sistema di controllo dell'integrità esegue un lavoro costante e le modifiche rapide e di grandi dimensioni non rappresentano una minaccia per la stabilità del sistema. Questo è in realtà il modo in cui Amazon Route 53 gestisce i controlli dell'integrità degli endpoint (come gli indirizzi IP) per stabilire come gli utenti finali vengono instradati verso di loro.

Livello di rischio associato se questa best practice non fosse adottata: basso

Guida all'implementazione

- Esegui un lavoro costante in modo che i sistemi non presentino guasti quando si verificano cambiamenti rapidi e significativi nel carico.
- Implementazione di dipendenze con accoppiamento debole Le dipendenze come sistemi di accodamento, sistemi di streaming, flussi di lavoro e bilanciatori del carico sono con accoppiamento debole. L'accoppiamento debole aiuta a isolare il comportamento di un componente dagli altri componenti che dipendono da esso, aumentando la resilienza e l'agilità.
  - The Amazon Builders' Library: Reliability, constant work, and a good cup of coffee

- AWS re:Invent 2018: Chiudere i circuiti e aprire le menti: come prendere il controllo dei sistemi, grandi e piccoli (include un lavoro costante) ARC337
  - Per l'esempio di un sistema di controllo dell'integrità che monitora 100.000 server, progetta i carichi di lavoro in modo che le dimensioni dei payload rimangano costanti indipendentemente dal numero di successi o di fallimenti.

Risorse

Documenti correlati:

- AmazonEC2: garantire l'idempotenza
- The Amazon Builders' Library: difficoltà dei sistemi distribuiti
- The Amazon Builders' Library: Reliability, constant work, and a good cup of coffee

Video correlati:

- AWS New York Summit 2019: introduzione alle architetture basate sugli eventi e ad Amazon (05) EventBridge MAD2
- AWS re:Invent 2018: Chiudere i circuiti e aprire le menti: come assumere il controllo di sistemi, grandi e piccoli (include un lavoro costante) ARC337
- AWS re:Invent 2018: Close Loops and Opening Minds: Come prendere il controllo dei sistemi, grandi e piccoli ARC337 (include accoppiamento libero, lavoro costante, stabilità statica)
- AWS re:Invent 2019: Passaggio ad architetture basate sugli eventi (08) SVS3

REL04-BP04 Rendi tutte le risposte idempotenti

Un servizio idempotente promette il completamento di ogni richiesta esattamente una volta, in modo tale che effettuare più richieste identiche abbia lo stesso effetto di effettuare una singola richiesta. Un servizio idempotente semplifica ad un client l'implementazione di nuovi tentativi senza temere che una richiesta venga elaborata erroneamente più volte. A tale scopo, i client possono inviare API richieste con un token di idempotenza: lo stesso token viene utilizzato ogni volta che la richiesta viene ripetuta. Un servizio idempotente API utilizza il token per restituire una risposta identica alla risposta restituita la prima volta che la richiesta è stata completata.

In un sistema distribuito, è facile eseguire un'operazione al massimo una volta (il client effettua una sola richiesta) o almeno una volta (la richiesta continua finché il client non ottiene la conferma

dell'esito positivo). Tuttavia, è difficile garantire che un'operazione sia idempotente, il che significa che viene eseguita esattamente una volta, in modo tale che effettuare più richieste identiche abbia lo stesso effetto di effettuare una singola richiesta. Utilizzando i token di idempotenza inAPIs, i servizi possono ricevere una richiesta mutante una o più volte senza creare record duplicati o effetti collaterali.

Livello di rischio associato se questa best practice non fosse adottata: medio

Guida all'implementazione

- Rendi tutte le risposte idempotenti. Un servizio idempotente promette il completamento di ogni richiesta esattamente una volta, in modo tale che effettuare più richieste identiche abbia lo stesso effetto di effettuare una singola richiesta.
  - I client possono inviare API richieste con un token di idempotenza: lo stesso token viene utilizzato ogni volta che la richiesta viene ripetuta. Un servizio idempotente API utilizza il token per restituire una risposta identica alla risposta restituita la prima volta che la richiesta è stata completata.
    - AmazonEC2: garantire l'idempotenza

Risorse

Documenti correlati:

- AmazonEC2: garantire l'idempotenza
- The Amazon Builders' Library: difficoltà dei sistemi distribuiti
- The Amazon Builders' Library: Reliability, constant work, and a good cup of coffee

Video correlati:

- AWS New York Summit 2019: introduzione alle architetture basate sugli eventi e ad Amazon (05) EventBridge MAD2
- AWS re:Invent 2018: Close Loops and Opening Minds: Come prendere il controllo di sistemi, grandi e piccoli ARC337 (include accoppiamento libero, lavoro costante, stabilità statica)
- AWS re:Invent 2019: Passaggio ad architetture basate sugli eventi (08) SVS3

REL5. Come si progettano le interazioni in un sistema distribuito per mitigare o affrontare gli errori?

I sistemi distribuiti si basano sulle reti di comunicazione per interconnettere i componenti (ad esempio server o servizi). Il carico di lavoro deve funzionare in modo affidabile nonostante la perdita o la latenza dei dati su queste reti. I componenti del sistema distribuito devono funzionare in modo da non influire negativamente su altri componenti o sul carico di lavoro. Queste best practice permettono ai carichi di lavoro di tollerare le sollecitazioni o i guasti, recuperare più rapidamente e mitigare l'impatto di tali problemi. Il risultato è un miglioramento del tempo medio di recupero (MTTR).

Best practice

- REL05-BP01 Implementa una degradazione graduale per trasformare le dipendenze rigide applicabili in dipendenze morbide
- REL05-BP02 Richieste Throttle
- REL05-BP03 Controlla e limita le chiamate di nuovo tentativo
- REL05-BP04 Fallisci velocemente e limita le code
- REL05-BP05 Imposta i timeout dei client
- REL05-BP06 Rendere i sistemi senza stato ove possibile
- REL05-BP07 Implementare leve di emergenza

REL05-BP01 Implementa una degradazione graduale per trasformare le dipendenze rigide applicabili in dipendenze morbide

I componenti dell'applicazione devono continuare a svolgere la loro funzione principale anche se le dipendenze non sono disponibili. Potrebbero fornire dati leggermente obsoleti, dati alternativi o addirittura nessun dato. Ciò garantisce che la funzionalità complessiva del sistema sia ostacolata solo in minima parte da errori localizzati, garantendo al contempo il valore aziendale intrinseco.

Risultato desiderato: quando le dipendenze di un componente non sono integre, il componente stesso può comunque funzionare, anche se in modo degradato. Le modalità di errore dei componenti devono essere considerate come funzionamenti normali. I flussi di lavoro devono essere progettati in modo tale che questi errori non conducano a un fallimento completo o almeno a stati prevedibili e recuperabili.

Anti-pattern comuni:

- Mancata identificazione della funzionalità aziendale di base necessaria. Mancata verifica del funzionamento dei componenti anche in caso di errori di dipendenza.
- Mancata restituzione di dati sugli errori o quando solo una delle dipendenze non è disponibile e possono comunque essere restituiti risultati parziali.
- Creazione di uno stato incoerente quando una transazione non va a buon fine parzialmente.
- Mancata disponibilità di alternative per accedere a un archivio di parametri centralizzato.
- Invalidare o svuotare lo stato locale a seguito di un aggiornamento non riuscito senza considerare le conseguenze di tale operazione.

Vantaggi dell'adozione di questa best practice: la normale riduzione delle prestazioni migliora la disponibilità del sistema nel suo complesso e conserva la funzionalità delle funzioni più importanti anche in caso di errori.

Livello di rischio associato se questa best practice non fosse adottata: elevato

Guida all'implementazione

L'implementazione di una normale riduzione delle prestazioni aiuta a ridurre al minimo l'impatto degli errori di dipendenza sul funzionamento dei componenti. Idealmente, un componente rileva gli errori nelle dipendenze e trova soluzioni alternative in modo da avere un impatto minimo sugli altri componenti o clienti.

Progettare per una normale riduzione delle prestazioni significa considerare le potenziali modalità di errore durante la progettazione delle dipendenze. Per ogni modalità di errore, disponi di un modo per fornire la maggior parte delle funzionalità o almeno quelle più critiche del componente a chiamanti o clienti. Queste considerazioni possono diventare requisiti aggiuntivi testabili e verificabili. Idealmente, un componente è in grado di svolgere la sua funzione principale in modo accettabile anche in caso di errore di una o più dipendenze.

Questa è una discussione di carattere tanto commerciale quanto tecnico. Tutti i requisiti aziendali sono importanti e devono essere soddisfatti, se possibile. Tuttavia, ha ancora senso chiedersi cosa dovrebbe succedere quando non tutti i requisiti possono essere soddisfatti. Un sistema può essere progettato per essere disponibile e coerente, ma nelle circostanze in cui è necessario eliminare un requisito, qual è quello più importante? Per l'elaborazione dei pagamenti, potrebbe essere la coerenza. Per un'applicazione in tempo reale, potrebbe essere la disponibilità. Per un sito Web rivolto ai clienti, la risposta può dipendere dalle aspettative dei clienti.

Il significato di ciò dipende dai requisiti del componente e da ciò che dovrebbe essere considerato la sua funzione principale. Per esempio:

- Un sito di e-commerce potrebbe visualizzare dati provenienti da più sistemi diversi, come consigli personalizzati, prodotti con il punteggio più alto e lo stato degli ordini dei clienti sulla pagina di destinazione. Quando in un sistema upstream si verifica un errore, ha comunque senso mostrare tutto il resto, invece di mostrare una pagina di errore a un cliente.
- Un componente che esegue la scrittura in batch può continuare a elaborare un batch se una delle singole operazioni fallisce. Dovrebbe essere semplice implementare un meccanismo di ripetizione dei tentativi. A tale scopo, è sufficiente restituire al chiamante informazioni su quali operazioni hanno avuto successo, quali e perché non sono riuscite, oppure inserendo le richieste non riuscite in una coda DLQ per implementare nuovi tentativi asincroni. Anche le informazioni sulle operazioni non riuscite devono essere registrate.
- Un sistema che elabora le transazioni deve verificare che vengano eseguiti tutti gli aggiornamenti o nessun aggiornamento. Per le transazioni distribuite, il modello Saga può essere utilizzato per ripristinare le operazioni precedenti nel caso in cui fallisca un'operazione successiva della stessa transazione. Qui, la funzione principale è mantenere la coerenza.
- I sistemi critici dal punto di vista temporale dovrebbero essere in grado di gestire le dipendenze che non rispondono in modo tempestivo. In questi casi, è possibile utilizzare lo schema dell'interruttore. Quando inizia a verificarsi il timeout delle risposte di una dipendenza, il sistema può passare a uno stato chiuso in cui non vengono effettuate chiamate aggiuntive.
- Un'applicazione può leggere i parametri da un archivio di parametri. Può essere utile creare immagini di container con un set di parametri predefinito e utilizzarli nel caso in cui l'archivio dei parametri non sia disponibile.

Si noti che i percorsi seguiti in caso di errore dei componenti devono essere testati e devono essere significativamente più semplici del percorso primario. In genere, è consigliabile evitare strategie di fallback.

Passaggi dell'implementazione

Identifica le dipendenze esterne e interne. Considera i tipi di errore che si possono verificare nelle dipendenze. Considera i modi per ridurre al minimo l'impatto negativo sui sistemi upstream e downstream e sui clienti durante questi errori.

Di seguito è riportato un elenco di dipendenze e di come ridurre normalmente le prestazioni quando si verifica un errore a livello di dipendenze:

1. Errore parziale delle dipendenze: un componente può effettuare più richieste ai sistemi downstream, sia come richieste multiple a un sistema sia come richiesta a più sistemi. A seconda del contesto aziendale, possono essere appropriate diverse modalità di gestione (per maggiori dettagli, consulta gli esempi precedenti nella Guida all'implementazione).
2. Un sistema downstream non è in grado di elaborare le richieste a causa del carico elevato: se le richieste rivolte a un sistema downstream non vanno costantemente a buon fine, non ha senso continuare a riprovare. Ciò può creare un carico aggiuntivo su un sistema già sovraccarico e rendere più difficile il ripristino. Qui è possibile utilizzare lo schema dell'interruttore, che monitora le chiamate non riuscite a un sistema downstream. Se un numero elevato di chiamate ha esito negativo, interromperà l'invio di altre richieste al sistema downstream e solo occasionalmente lascerà passare le chiamate per verificare se il sistema downstream è nuovamente disponibile.
3. Un archivio di parametri non è disponibile: per trasformare un archivio di parametri, è possibile utilizzare la cache delle dipendenze a protezione debole o i valori predefiniti integri inclusi nelle immagini del container o del computer. Nota che queste impostazioni predefinite devono essere mantenute e incluse nelle suite di test. up-to-date
4. Un servizio di monitoraggio o altra dipendenza non funzionale non è disponibile: se un componente non è in grado di inviare a intermittenza log, metriche o tracce a un servizio di monitoraggio centralizzato, spesso è meglio continuare a eseguire le funzioni aziendali come al solito. Non registrare o eseguire il push delle metriche in modo invisibile all'utente per un lungo periodo di tempo spesso non è una procedura accettabile. Inoltre, in alcuni casi d'uso potrebbero essere necessari dati di controllo completi per soddisfare i requisiti di conformità.
5. Un'istanza primaria di un database relazionale potrebbe non essere disponibile: come quasi tutti i database relazionali, Amazon Relational Database Service può presentare solo un'istanza di scrittura primaria. Questo crea un unico punto di errore per i carichi di lavoro di scrittura e rende più difficile il dimensionamento. Questo problema può essere parzialmente mitigato utilizzando una configurazione Multi-AZ per una disponibilità elevata o Amazon Aurora Serverless per un migliore dimensionamento. Per requisiti di disponibilità molto elevati, può essere logico non fare affatto affidamento sull'istanza di scrittura primaria. Per le query che si limitano a leggere, è possibile utilizzare repliche di lettura, che forniscono ridondanza e la possibilità di aumentare orizzontalmente e anche verticalmente. Le operazioni di scrittura possono essere memorizzate nel buffer, ad esempio in una coda Amazon Simple Queue Service, in modo che le richieste di scrittura dei clienti possano comunque essere accettate anche se l'istanza primaria non è temporaneamente disponibile.

Risorse

Documenti correlati:

- Amazon API Gateway: limita le API richieste per un throughput migliore
- CircuitBreaker(riassume Circuit Breaker tratto da «Release It!» libro)
- Tentativi di errore e backoff esponenziale in AWS
- Michael Nygard "Release It! Design and Deploy Production-Ready Software"
- The Amazon Builders' Library: evitare il fallback nei sistemi distribuiti
- The Amazon Builders' Library: evitare insormontabili backlog di code
- The Amazon Builders' Library: sfide e strategie del caching
- The Amazon Builders' Library: timeout, nuovi tentativi e backoff con jitter

Video correlati:

- Riprova, backoff e jitter: AWS re:Invent 2019: Presentazione di The Amazon Builders' Library () DOP328

Esempi correlati:

- Well-Architected Lab (Livello 300): Implementing Health Checks and Managing Dependencies to Improve Reliability

REL05-BP02 Richieste Throttle

Usa le richieste di limitazione (della larghezza di banda della rete) per mitigare l'esaurimento delle risorse dovuto ad aumenti imprevisti della domanda. Le richieste inferiori ai tassi di limitazione vengono elaborate, mentre quelle che superano il limite definito vengono rifiutate con un messaggio di risposta che indica che la richiesta è stata limitata.

Risultato desiderato: i picchi di volume di grandi dimensioni dovuti a improvvisi aumenti del traffico dei clienti, attacchi di flooding o tempeste di ripetizioni dei tentativi sono mitigati dalla limitazione (della larghezza di banda della rete) delle richieste, che consente ai carichi di lavoro di continuare la normale elaborazione del volume di richieste supportato.

Anti-pattern comuni:

- APIi throttles degli endpoint non vengono implementati o vengono lasciati ai valori predefiniti senza considerare i volumi previsti.
- APIgli endpoint non sono testati in termini di carico o i limiti di throttling non sono testati.
- Limitazione (della larghezza di banda della rete) dei tassi di richiesta senza considerare le dimensioni o la complessità delle richieste.
- Verifica delle percentuali massime di richieste o delle dimensioni massime delle richieste, senza però testarle congiuntamente.
- Le risorse non vengono allocate entro gli stessi limiti stabiliti durante i test.
- I piani di utilizzo non sono stati configurati o presi in considerazione per gli utenti da applicazione ad applicazione (A2A). API
- Gli utenti di code con scalabilità orizzontale non hanno configurato le impostazioni di simultaneità massima.
- La limitazione della velocità per indirizzo IP non è stata implementata.

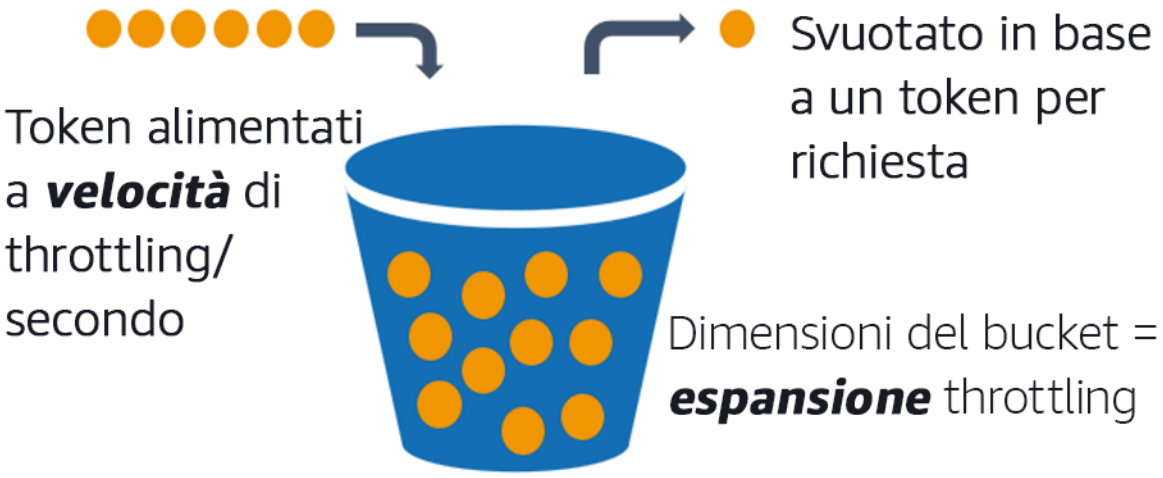
Vantaggi dell'adozione di questa best practice: i carichi di lavoro che stabiliscono limiti di accelerazione sono in grado di funzionare normalmente ed elaborare correttamente il caricamento delle richieste accettate in presenza di picchi di volume imprevisti. I picchi improvvisi o prolungati di richieste APIs e code vengono limitati e non esauriscono le risorse di elaborazione delle richieste. I limiti di velocità limitano i singoli richiedenti, in modo che volumi elevati di traffico provenienti da un unico indirizzo IP o utente non esauriscano le risorse, con ripercussioni sugli altri utenti. API

Livello di rischio associato se questa best practice non fosse adottata: elevato

Guida all'implementazione

I servizi devono essere progettati per elaborare una capacità nota di richieste; tale capacità può essere stabilita mediante test di carico. Se le percentuali di arrivo delle richieste superano i limiti, la risposta appropriata segnala che una richiesta ha subito la limitazione (della larghezza di banda della rete). Ciò consente all'utente di gestire l'errore e riprovare in un secondo momento.

Quando il servizio richiede un'implementazione della limitazione (della larghezza di banda della rete), prendi in considerazione l'implementazione dell'algoritmo token bucket, in cui un token conta come una richiesta. I token vengono alimentati a una specifica velocità di limitazione (della larghezza di banda della rete) al secondo e svuotati in modo asincrono in base a un token per richiesta.



Algoritmo token bucket.

Amazon API Gateway implementa l'algoritmo token bucket in base ai limiti di account e regione e può essere configurato per cliente con piani di utilizzo. Inoltre, Amazon Simple Queue Service (AmazonSQS) e Amazon Kinesis possono bufferizzare le richieste per ridurre la frequenza delle richieste e consentire tassi di limitazione più elevati per le richieste che possono essere soddisfatte. Infine, puoi implementare la limitazione della velocità con utenti specifici che generano AWS WAFun carico insolitamente elevato. API

Passaggi dell'implementazione

È possibile configurare API Gateway con limiti di limitazione APIs e restituire `429 Too Many Requests` errori quando i limiti vengono superati. Puoi utilizzarlo AWS WAF con i tuoi endpoint AWS AppSync e API Gateway per abilitare la limitazione della velocità in base all'indirizzo IP. Inoltre, laddove il sistema può tollerare l'elaborazione asincrona, è possibile inserire i messaggi in una coda o in un flusso per velocizzare le risposte ai client del servizio, il che consente di aumentare i tassi di limitazione (della larghezza di banda della rete).

Con l'elaborazione asincrona, dopo aver configurato Amazon SQS come fonte di eventi per AWS Lambda, puoi configurare la massima concorrenza per evitare che tassi di eventi elevati consumino la quota di esecuzione simultanea dell'account disponibile necessaria per altri servizi del tuo carico di lavoro o account.

Sebbene API Gateway fornisca un'implementazione gestita del token bucket, nei casi in cui non è possibile utilizzare API Gateway, è possibile sfruttare le implementazioni open source specifiche del linguaggio (vedere gli esempi correlati in Risorse) del token bucket per i propri servizi.

- Comprendi e configura i limiti di limitazione di API Gateway a livello di account per regione, API per fase e API a livello di piano di utilizzo in base alla chiave.
- Applica le regole AWS WAF di limitazione della velocità a API Gateway e agli AWS AppSync endpoint per proteggerti dalle inondazioni e bloccare i malintenzionati. IPs Le regole di limitazione della velocità possono essere configurate anche sulle AWS AppSync API chiavi per i consumatori A2A.
- Valuta se hai bisogno di un maggiore controllo della limitazione della velocità rispetto alla limitazione della velocità e AWS AppSync APIs, in tal caso, configura un API Gateway davanti al tuo endpoint. AWS AppSync
- Quando le SQS code Amazon sono configurate come trigger per i consumatori di code Lambda, imposta la massima concorrenza su un valore che elabora abbastanza da soddisfare i tuoi obiettivi di livello di servizio ma non consuma limiti di concorrenza che influiscono su altre funzioni Lambda. Valuta la possibilità di impostare la simultaneità riservata su altre funzioni Lambda nello stesso account e nella stessa regione quando utilizzi le code con Lambda.
- Usa API Gateway con integrazioni di servizi native con Amazon SQS o Kinesis per bufferizzare le richieste.
- Se non puoi usare API Gateway, consulta le librerie specifiche del linguaggio per implementare l'algoritmo token bucket per il tuo carico di lavoro. Controlla la sezione degli esempi e cerca una libreria adatta.
- Verifica i limiti che intendi impostare o che prevedi di incrementare e documenta i limiti testati.
- Non aumentare i limiti oltre i valori stabiliti durante i test. Quando si aumenta un limite, verifica che le risorse allocate siano equivalenti o superiori a quelle degli scenari di test prima di applicare l'aumento.

Risorse

Best practice correlate:

- REL04-BP03 Fai un lavoro costante
- REL05-BP03 Controlla e limita le chiamate di nuovo tentativo

Documenti correlati:

- Amazon API Gateway: limita le API richieste per un throughput migliore
- AWS WAF: Rate-based rule statement

- Introduzione della massima concorrenza nell' AWS Lambda utilizzo di Amazon SQS come fonte di eventi
- AWS Lambda: Maximum Concurrency

Esempi correlati:

- Le tre regole più importanti basate sulle AWS WAF tariffe
- Java Bucket4j
- Algoritmo token bucket per Python
- Algoritmo token bucket a livello di nodo
- . NETLimitazione della velocità di threading del sistema

Video correlati:

- Implementazione delle migliori pratiche di API sicurezza GraphQL con AWS AppSync

Strumenti correlati:

- Amazon API Gateway
- AWS AppSync
- Amazon SQS
- Amazon Kinesis
- AWS WAF

REL05-BP03 Controlla e limita le chiamate di nuovo tentativo

Utilizza il backoff esponenziale per rieseguire le richieste a intervalli progressivamente più lunghi tra i singoli nuovi tentativi. Introduci il jitter tra i tentativi per la randomizzazione degli intervalli di ripetizione. Limita il numero massimo di tentativi.

Risultato desiderato: i componenti tipici di un sistema software distribuito includono server, sistemi di bilanciamento del carico, database e server. DNS Durante il normale funzionamento, questi componenti possono rispondere alle richieste con errori temporanei o limitati e anche errori che sarebbero persistenti indipendentemente dai nuovi tentativi. Quando i client effettuano richieste ai servizi, le richieste consumano risorse tra cui memoria, thread, connessioni, porte o altre risorse

limitate. Controllare e limitare i nuovi tentativi è una strategia per liberare risorse e ridurne al minimo il consumo in modo che i componenti del sistema sottoposti a stress non vengano sovraccaricati.

Quando vanno in timeout o ricevono risposte di errore, le richieste client devono decidere se eseguire o meno nuovi tentativi. Se vengono eseguiti nuovi tentativi, questi verranno eseguiti con un backoff esponenziale con jitter e un numero massimo di tentativi. Di conseguenza, i servizi e i processi backend riducono il carico e i tempi di riparazione automatica, con un conseguente ripristino più rapido e una corretta gestione delle richieste.

Anti-pattern comuni:

- Implementazione di nuovi tentativi senza aggiungere il backoff esponenziale, il jitter e il numero massimo di tentativi. Il backoff e il jitter aiutano a evitare picchi di traffico artificiali dovuti a tentativi involontariamente coordinati a intervalli standard.
- Implementazione di nuovi tentativi senza testarne gli effetti o presupponendo che i nuovi tentativi siano già integrati in scenari di ripetizione e senza test. SDK
- La mancata comprensione dei codici di errore pubblicati nelle dipendenze porta a ritentare tutti gli errori, compresi quelli la cui causa è chiara e indica la mancanza di autorizzazione, un errore di configurazione o un'altra condizione che prevedibilmente non si risolverà senza un intervento manuale.
- Mancata gestione delle best practice di osservabilità, compresi il monitoraggio e la segnalazione di ripetuti guasti del servizio, in modo che i problemi sottostanti siano resi noti e possano essere risolti.
- Sviluppo di meccanismi di ripetizione personalizzati quando le funzionalità di ripetizione dei tentativi integrate o di terze parti sono sufficienti.
- Riprovare su più livelli dello stack di applicazioni in modo da accrescere in modo significativo i nuovi tentativi e pertanto da consumare ulteriormente le risorse in una tempesta di ripetizioni dei tentativi. Assicurati di comprendere in che modo questi errori influiscono sulla tua applicazione e sulle dipendenze su cui fai affidamento, quindi implementa i nuovi tentativi a un solo livello.
- Riesecuzione delle chiamate dei servizi non idempotenti, con effetti collaterali imprevisti come risultati duplicati.

Vantaggi dell'adozione di questa best practice: i nuovi tentativi aiutano i client a ottenere i risultati desiderati quando le richieste non vanno a buon fine, ma consumano più tempo del server per ottenere le risposte corrette desiderate. Quando gli errori sono rari o transitori, i nuovi tentativi funzionano correttamente. Quando gli errori sono causati da un sovraccarico di risorse, i nuovi

tentativi possono peggiorare le cose. L'aggiunta di un backoff esponenziale con jitter ai tentativi dei client consente ai server di recuperare risorse quando gli errori sono causati dal sovraccarico delle risorse. Il jitter evita l'allineamento delle richieste in picchi e il backoff riduce l'aumento del carico causato dall'aggiunta di nuovi tentativi al normale carico delle richieste. Infine, è importante configurare un numero massimo di tentativi o il tempo trascorso per evitare la creazione di backlog che producono errori metastabili.

Livello di rischio associato se questa best practice non fosse adottata: elevato

Guida all'implementazione

Controlla e limita le chiamate riproposte. Utilizza il backoff esponenziale per eseguire nuovi tentativi dopo intervalli progressivamente più lunghi. Introduci il jitter per la randomizzazione degli intervalli di ripetizione e limitare il numero massimo di tentativi.

Per impostazione predefinita, alcuni AWS SDKs implementano nuovi tentativi e il backoff esponenziale. Utilizza queste AWS implementazioni integrate laddove applicabile nel tuo carico di lavoro. Implementa una logica simile nel tuo carico di lavoro nelle chiamate di servizi idempotenti e i cui tentativi migliorano la disponibilità dei client. Potrai decidere quali sono i timeout e quando cessare i tentativi in base al tuo caso d'uso. Crea ed esegui scenari di test per quei casi d'uso relativi ai nuovi tentativi.

Passaggi dell'implementazione

- Determina il livello ottimale nello stack di applicazioni per implementare nuovi tentativi per i servizi su cui si basa l'applicazione.
- Sii consapevole delle strategie esistenti SDKs che implementano strategie di ripetizione comprovate con backoff e jitter esponenziali per il linguaggio che preferisci, e preferiscile alla stesura di implementazioni personalizzate con nuovi tentativi.
- Verifica che i servizi siano caratterizzati dall'idempotenza prima dell'implementazione di nuovi tentativi. Una volta implementati i nuovi tentativi, assicurati che siano testati e che vengano regolarmente eseguiti in produzione.
- Quando chiami il AWS servizioAPIs, usa AWS SDKsand e AWS CLIcomprendi le opzioni di configurazione Retry. Determina se le impostazioni predefinite sono adatte al tuo caso d'uso, esegui i test e regola i valori secondo necessità.

Risorse

Best practice correlate:

- REL04-BP04 Rendi tutte le risposte idempotenti
- REL05-BP02 Richieste Throttle
- REL05-BP04 Fallisci velocemente e limita le code
- REL05-BP05 Imposta i timeout dei client
- REL11-BP01 Monitora tutti i componenti del carico di lavoro per rilevare i guasti

Documenti correlati:

- Errore, tentativi e backoff esponenziale in AWS
- The Amazon Builders' Library: timeout, nuovi tentativi e backoff con jitter
- Exponential Backoff and Jitter
- Rendere sicuri i nuovi tentativi con idempotent APIs

Esempi correlati:

- Spring Retry
- Resilience4j Retry

Video correlati:

- Riprova, backoff e jitter: AWS re:Invent 2019: Presentazione di The Amazon Builders' Library () DOP328

Strumenti correlati:

- AWS SDKsStrumenti e strumenti: Riprova il comportamento
- AWS Command Line Interface: AWS CLI riprova

REL05-BP04 Fallisci velocemente e limita le code

Se un servizio non è in grado di rispondere correttamente a una richiesta, procede ad anticipare l'errore (fail fast). Ciò consente il rilascio delle risorse associate a una richiesta e permette al servizio di recuperare le risorse se queste sono in esaurimento. L'anticipazione degli errori (fail fast) è un modello di progettazione software consolidato che può essere usato per creare carichi di lavoro altamente affidabili nel cloud. Anche l'accodamento è un modello di integrazione

aziendale consolidato che può semplificare il carico e consentire ai client di rilasciare risorse quando l'elaborazione asincrona può essere tollerata. Quando un servizio è in grado di rispondere correttamente in condizioni normali ma restituisce un esito negativo quando la frequenza delle richieste è troppo alta, utilizza una coda per memorizzare le richieste nel buffer. Tuttavia, non consentire la creazione di backlog di code lunghe che possono comportare l'elaborazione di richieste obsolete già dismesse dal client.

Risultato desiderato: quando i sistemi rilevano conflitti a livello di risorse, timeout, eccezioni o errori che rendono irraggiungibili gli obiettivi dei livelli di servizio, le strategie volte ad anticipare l'errore (fail fast) consentono un ripristino più rapido del sistema. I sistemi che devono assorbire i picchi di traffico e sono in grado di gestire l'elaborazione asincrona possono migliorare l'affidabilità consentendo ai client di rilasciare rapidamente le richieste utilizzando le code per archiviare le richieste nei servizi di backend. Quando le richieste vengono memorizzate nei buffer delle code, vengono implementate strategie di gestione delle code per evitare backlog ingestibili.

Anti-pattern comuni:

- Implementazione delle code di messaggi ma non configurazione delle code di lettere morte (DLQ) o degli allarmi sui DLQ volumi per rilevare quando un sistema è in errore.
- Mancata misurazione dell'età dei messaggi in una coda, misurazione della latenza per capire quando gli utenti della coda sono in ritardo o generano errori che causano un nuovo tentativo.
- Mancata cancellazione dei messaggi nel backlog da una coda quando non è più necessario elaborare questi messaggi se l'azienda non lo richiede più.
- La configurazione delle code first in first out (FIFO) quando last in first out (LIFO) risponderebbe meglio alle esigenze dei clienti, ad esempio quando non è richiesto un ordine rigoroso e l'elaborazione degli arretrati ritarda tutte le richieste nuove e urgenti, con conseguenti violazioni dei livelli di servizio per tutti i clienti.
- Esporre le code interne ai clienti anziché esporre i clienti a questi ultimi, gestire l'assunzione di lavoro e inserire le APIs richieste nelle code interne.
- Combinazione di un numero eccessivo di tipi di richieste di lavoro in un'unica coda: ciò può aggravare le condizioni dei backlog in seguito alla distribuzione delle richieste di risorse tra i tipi di richiesta.
- Elaborazione di richieste complesse e semplici nella stessa coda, nonostante siano necessari monitoraggio, timeout e allocazioni di risorse diversi.

- Mancata convalida degli input o utilizzo di asserzioni per implementare meccanismi in grado di anticipare l'errore (fail fast) nel software che generano eccezioni a componenti di livello superiore in grado di gestire normalmente gli errori.
- Mancata rimozione delle risorse in errore dall'instradamento delle richieste, soprattutto quando gli errori generano risultati sia positivi che negativi dovuti ad arresti anomali e riavvii, errori intermittenti a livello di dipendenze, capacità ridotta o perdita di pacchetti di rete.

Vantaggi dell'adozione di questa best practice: i sistemi in grado di anticipare l'errore (fail fast) sono più facili da sottoporre al debug e alla correzione degli errori e spesso presentano problemi di codifica e configurazione prima che le versioni vengano pubblicate in produzione. I sistemi che incorporano strategie di accodamento efficaci forniscono maggiore resilienza e affidabilità in caso di picchi di traffico e di condizioni intermittenti di errore del sistema.

Livello di rischio associato se questa best practice non fosse adottata: elevato

Guida all'implementazione

Le strategie volte ad anticipare l’errore (fail fast) possono essere codificate in soluzioni software e configurate nell'infrastruttura. Oltre ad anticipare l'errore (fail fast), le code sono una tecnica semplice ma affidabile di definizione dell'architettura che consente il caricamento senza problemi di componenti disaccoppiati del sistema. Amazon CloudWatch offre funzionalità di monitoraggio e allarme in caso di guasti. Una volta accertato il malfunzionamento di un sistema, è possibile richiamare strategie di mitigazione, ad esempio per evitare problemi dovuti a risorse danneggiate. Quando i sistemi implementano code con Amazon SQS e altre tecnologie di coda per semplificare il caricamento, devono considerare come gestire gli arretrati delle code e gli errori di consumo dei messaggi.

Passaggi dell'implementazione

- Implementa asserzioni programmatiche o metriche specifiche nel tuo software e usale per ricevere avvisi espliciti in caso di problemi di sistema. Amazon ti CloudWatch aiuta a creare metriche e allarmi basati sullo schema e SDK sulla strumentazione dei log delle applicazioni.
- Usa CloudWatch metriche e allarmi per evitare che risorse compromettano la velocità di elaborazione o che impediscono ripetutamente di elaborare le richieste.
- Utilizza l'elaborazione asincrona progettando APIs per accettare richieste e aggiungere richieste alle code interne utilizzando AmazonSQS, quindi rispondi al client che produce i messaggi con un messaggio di successo in modo che il client possa liberare risorse e passare ad altre attività mentre gli utenti della coda di backend elaborano le richieste.

- Misura e monitora la latenza di elaborazione delle code generando una CloudWatch metrica ogni volta che togli un messaggio dalla coda confrontando «now» con il timestamp del messaggio.
- Quando gli errori impediscono la corretta elaborazione dei messaggi o il traffico aumenta a livelli tali da impedirne l'elaborazione in base agli accordi sul livello di servizio, escludi il traffico obsoleto o in eccesso indirizzandolo a una coda per il traffico eccedente. Ciò consente l'elaborazione prioritaria del nuovo processo e del processo più vecchio quando si rende disponibile nuova capacità. Questa tecnica è un'approssimazione dell'LIFOelaborazione e consente la normale elaborazione del sistema per tutti i nuovi lavori.
- Usa le code DLQ o le code di reindirizzamento per spostare i messaggi che non possono essere elaborati dal backlog in una posizione che può essere ricercata e risolta in un secondo momento.
- Riprova o, se possibile, elimina i vecchi messaggi confrontandoli con il timestamp del messaggio ed eliminando i messaggi che non sono più rilevanti per il client richiedente.

Risorse

Best practice correlate:

- REL04-BP02 Implementare dipendenze liberamente accoppiate
- REL05-BP02 Richieste Throttle
- REL05-BP03 Controlla e limita le chiamate di nuovo tentativo
- REL06-BP02 Definire e calcolare le metriche (aggregazione)
- REL06-BP07 Monitora il end-to-end tracciamento delle richieste attraverso il sistema

Documenti correlati:

- Evitare insormontabili backlog di code
- Fail Fast
- Come posso evitare un crescente arretrato di messaggi nella mia SQS coda Amazon?
- Elastic Load Balancing: spostamento zonale
- Amazon Application Recovery Controller: controllo del routing per il failover del traffico

Esempi correlati:

- Modelli di integrazione aziendale: canale DLQ

Video correlati:

- AWS re:Invent 2022 - Utilizzo di applicazioni Multi-AZ ad alta disponibilità

Strumenti correlati:

- Amazon SQS
- Amazon MQ
- AWS IoT Core
- Amazon CloudWatch

REL05-BP05 Imposta i timeout dei client

Imposta i timeout in modo appropriato per connessioni e richieste, verificali sistematicamente e non fare affidamento sui valori predefiniti perché non fanno riferimento alle specifiche del carico di lavoro.

Risultato desiderato: i timeout dei client devono considerare il costo per client, server e carico di lavoro associato all'attesa di richieste il cui completamento richiede una quantità di tempo anomala. Poiché non è possibile conoscere la causa esatta di un timeout, i client devono fare riferimento ai servizi per sviluppare ipotesi sulle cause probabili e sui timeout appropriati.

Il timeout delle connessioni client si verifica in base ai valori configurati. Dopo aver rilevato un timeout, i client decidono di riprovare o aprire un interruttore. Questi modelli evitano la generazione di richieste che potrebbero aggravare una condizione di errore sottostante.

Anti-pattern comuni:

- Non essere a conoscenza dei timeout di sistema o dei timeout predefiniti.
- Non essere a conoscenza dei normali tempi di completamento delle richieste.
- Non essere a conoscenza delle possibili cause dei tempi anomali necessari per il completamento delle richieste o dei costi in termini di prestazioni di client, servizio o carico di lavoro associati all'attesa di tali completamenti.
- Non essere consapevoli della probabilità che una rete danneggiata causi un errore di esecuzione della richiesta solo al raggiungimento del timeout, nonché dei costi in termini di prestazioni del client e del carico di lavoro derivanti dalla mancata adozione di un timeout più breve.
- Non testare gli scenari di timeout sia per le connessioni che per le richieste.

- Impostazione di timeout troppo elevati, che può comportare lunghi tempi di attesa e aumentare l'utilizzo delle risorse.
- Impostazione di timeout troppo bassi, con conseguenti errori artificiali.
- Mancata verifica degli schemi per gestire gli errori di timeout per chiamate remote come interruttori e nuovi tentativi.
- Non considerare il monitoraggio delle percentuali di errore delle chiamate dei servizi, degli obiettivi del livello di servizio per la latenza e dei valori anomali della latenza. Queste metriche possono fornire informazioni sui timeout restrittivi o permissivi.

Vantaggi dell'adozione di questa best practice: i timeout delle chiamate remote sono configurati e i sistemi sono progettati per gestirli correttamente, in modo da preservare le risorse quando le chiamate remote rispondono in modo eccessivamente lento e gli errori di timeout vengono gestiti correttamente dai client di servizio.

Livello di rischio associato se questa best practice non fosse adottata: elevato

Guida all'implementazione

Imposta sia un timeout di connessione che un timeout della richiesta su qualsiasi chiamata della dipendenza del servizio e, generalmente, su qualsiasi chiamata tra i processi. Molti framework offrono funzionalità di timeout integrate, ma è necessario prestare attenzione perché alcuni sono caratterizzati da valori predefiniti infiniti o superiori a quelli accettabili per gli obiettivi dei tuoi servizi. Un valore troppo elevato riduce l'utilità del timeout perché le risorse continuano a essere consumate mentre il client attende che si verifichi il timeout. Un valore troppo basso può generare un aumento del traffico sul backend e una maggiore latenza perché vengono ritentate troppe richieste. In alcuni casi, questo può portare a interruzioni vere e proprie perché tutte le richieste vengono ritentate.

Considera quanto segue per determinare le strategie di timeout:

- L'elaborazione delle richieste può richiedere più tempo del normale a causa del loro contenuto, di problemi nel servizio di destinazione o di un errore nella partizione della rete.
- Le richieste con contenuti troppo costosi potrebbero consumare risorse server e client non necessarie. In questo caso, forzare il timeout di queste richieste e non eseguire nuovi tentativi possono preservare le risorse. I servizi dovrebbero, inoltre, proteggersi da contenuti eccessivamente costosi con limitazione (della larghezza di banda della rete) e timeout lato server.
- Per le richieste con tempi di elaborazione eccessivamente lunghi a causa di un'interruzione del servizio è possibile forzare il timeout e, quindi, eseguire un nuovo tentativo. È necessario

considerare i costi del servizio per la richiesta e il nuovo tentativo, ma se la causa è un problema localizzato, è probabile che un nuovo tentativo non sia costoso e riduca il consumo di risorse del client. Il timeout può anche liberare risorse del server a seconda della natura del problema.

- Per le richieste il cui completamento richiede troppo tempo o per risposte non distribuite dalla rete è possibile forzare il timeout e, quindi, eseguire un nuovo tentativo. Poiché la richiesta o la risposta non è stata distribuita, viene comunque restituito un errore indipendentemente dalla durata del timeout. Il timeout in questo caso non rilascerà le risorse del server, ma le risorse del client, con il conseguente miglioramento delle prestazioni del carico di lavoro.

Sfrutta gli schemi di progettazione consolidati, come i tentativi e gli interruttori automatici, per gestire i timeout in modo corretto e supportare approcci di tipo fail-fast. AWS SDKse AWS CLIconsentono la configurazione dei timeout di connessione e di richiesta e i nuovi tentativi con backoff e jitter esponenziali. AWS Lambdale funzioni supportano la configurazione dei timeout e, con AWS Step Functions, è possibile creare interruttori automatici a basso codice che sfruttano le integrazioni predefinite con i servizi e. AWS SDKs AWS App Mesh Envoy fornisce funzionalità di tipo timeout e interruttore.

Passaggi dell'implementazione

- Configura i timeout per le chiamate remote dei servizi e sfrutta le funzionalità di timeout integrate o le librerie di timeout open source.
- Quando il tuo carico di lavoro effettua chiamate con un AWS SDK, consulta la documentazione per la configurazione del timeout specifica della lingua.
  - Python
  - PHP
  - .NET
  - Ruby
  - Java
  - Go
  - Node.js
  - C++
- Quando utilizzi AWS CLI i comandi AWS SDKs or nel tuo carico di lavoro, configura i valori di timeout predefiniti impostando i valori di AWS configurazione predefiniti per e. `connectTimeoutInMillis` `tlsNegotiationTimeoutInMillis`

- Applica le opzioni della riga di comando `cli-connect-timeout` e controlla comandi singoli `cli-read-timeout` AWS CLI ai servizi. AWS
- Monitora le chiamate remote dei servizi per i timeout e imposta gli allarmi sugli errori persistenti in modo da poter gestire in modo proattivo gli scenari di errore.
- Implementa le CloudWatch metriche e il rilevamento delle CloudWatch anomalie sui tassi di errore delle chiamate, sugli obiettivi dei livelli di servizio per la latenza e sui valori anomali di latenza per fornire informazioni sulla gestione di timeout eccessivamente aggressivi o permissivi.
- Configura i timeout sulle funzioni Lambda.
- APII client Gateway devono implementare i propri tentativi quando gestiscono i timeout. APIGateway supporta un timeout di integrazione da 50 millisecondi a 29 secondi per le integrazioni downstream e non riprova quando le richieste di integrazione scade.
- Implementa il modello dell'interruttore per evitare di effettuare chiamate remote quando si è verificato il timeout. Apri l'interruttore per evitare chiamate non riuscite e chiudi l'interruttore quando le chiamate rispondono normalmente.
- Per i carichi di lavoro basati su container, esamina le funzionalità di App Mesh Envoy per sfruttare timeout e interruttori integrati.
- Utilizzali AWS Step Functions per creare interruttori automatici a basso codice per chiamate di assistenza remota, in particolare quando richiami integrazioni Step Functions AWS native SDKs e supportate per semplificare il carico di lavoro.

Risorse

Best practice correlate:

- REL05-BP03 Controlla e limita le chiamate di nuovo tentativo
- REL05-BP04 Fallisci velocemente e limita le code
- REL06-BP07 Monitora il end-to-end tracciamento delle richieste attraverso il sistema

Documenti correlati:

- AWS SDK: Tentativi e timeout
- The Amazon Builders' Library: timeout, nuovi tentativi e backoff con jitter
- Quote e note importanti di Amazon API Gateway
- AWS Command Line Interface: opzioni della riga di comando

- AWS SDK for Java 2.x: Configura API i timeout
- AWS Botocore utilizzando l'oggetto config e Config Reference
- AWS SDK for .NET: nuovi tentativi e timeout
- AWS Lambda: configurazione delle opzioni della funzione Lambda

Esempi correlati:

- Utilizzo del pattern di interruttore automatico con AWS Step Functions e Amazon DynamoDB
- Martin Fowler: CircuitBreaker

Strumenti correlati:

- AWS SDKs
- AWS Lambda
- Amazon SQS
- AWS Step Functions
- AWS Command Line Interface

REL05-BP06 Rendere i sistemi senza stato ove possibile

I sistemi non devono richiedere lo stato né eseguire l'offload dello stato in modo tale che, tra diverse richieste client, non vi sia alcuna dipendenza dai dati archiviati localmente su disco o in memoria. I server possono così essere sostituiti a piacimento senza compromettere la disponibilità.

Quando gli utenti o i servizi interagiscono con un'applicazione, spesso eseguono una serie di interazioni che formano una sessione. Una sessione è un dato univoco per gli utenti che persistono tra le richieste mentre utilizzano l'applicazione. Un'applicazione stateless è un'applicazione che non richiede la conoscenza delle interazioni precedenti e non memorizza le informazioni sulla sessione.

Una volta progettati per essere stateless, è possibile utilizzare servizi di elaborazione serverless, come o. AWS Lambda AWS Fargate

Oltre alla sostituzione dei server, un altro vantaggio delle applicazioni stateless è la possibilità di scalare orizzontalmente perché tutte le risorse di elaborazione disponibili (come EC2 istanze e AWS Lambda funzioni) possono soddisfare qualsiasi richiesta.

Vantaggi dell'adozione di questa best practice: i sistemi con progettazione stateless sono più adattabili alla scalabilità orizzontale, così da poter aggiungere o rimuovere capacità in base alle fluttuazioni di traffico e domanda. Sono inoltre intrinsecamente resistenti ai guasti e offrono flessibilità e agilità allo sviluppo delle applicazioni.

Livello di rischio associato se questa best practice non fosse adottata: medio

Guida all'implementazione

Trasforma le applicazioni in stateless. Le applicazioni stateless consentono la scalabilità orizzontale e sono tolleranti ai guasti di un singolo nodo. Analizza e individua i componenti della tua applicazione che mantengono lo stato dell'architettura. Questo processo ti aiuta a valutare il potenziale impatto della transizione a una progettazione stateless. Un'architettura stateless separa i dati degli utenti ed esegue l'offload dei dati della sessione, offrendo la flessibilità necessaria per scalare ogni componente in modo indipendente al fine di soddisfare le diverse richieste del carico di lavoro e ottimizzare l'utilizzo delle risorse.

Passaggi dell'implementazione

- Individua e comprendi i componenti stateful dell'applicazione.
- Suddividi i dati, separando e gestendo i dati dell'utente dalla logica dell'applicazione principale.
  - Amazon Cognito è in grado di separare i dati degli utenti dal codice dell'applicazione mediante funzionalità come pool di identità, pool di utenti e Amazon Cognito Sync.
  - Puoi separare dati degli utenti con AWS Secrets Manager, archiviando i segreti in un luogo sicuro e centralizzato. Il codice dell'applicazione pertanto non dovrà più memorizzare i segreti, rendendolo più sicuro.
  - Per archiviare dati non strutturati di grandi dimensioni, come immagini e documenti, prendi in considerazione l'utilizzo di Amazon S3. L'applicazione può recuperare questi dati quando richiesto, eliminando la necessità di archiviarli in memoria.
  - Utilizza Amazon DynamoDB per archiviare informazioni come i profili utente. L'applicazione può eseguire query su questi dati pressoché in tempo reale.
- Trasferisci i dati della sessione in un database, una cache o in file esterni.
  - Amazon ElastiCache, Amazon DynamoDB, Amazon Elastic File System (Amazon) e EFS Amazon MemoryDB sono esempi AWS di servizi che puoi utilizzare per scaricare i dati della sessione.
- Progetta un'architettura stateless dopo aver identificato lo stato e i dati dell'utente che devono essere mantenuti con la tua soluzione di archiviazione preferita.

Risorse

Best practice correlate:

- REL11-BP03 Automatizza la guarigione su tutti i livelli

Documenti correlati:

- The Amazon Builders' Library: evitare il fallback nei sistemi distribuiti
- The Amazon Builders' Library: evitare insormontabili backlog di code
- The Amazon Builders' Library: sfide e strategie del caching
- Best practice per Stateless Web Tier su AWS

REL05-BP07 Implementare leve di emergenza

Le leve di emergenza sono processi rapidi che possono mitigare l'impatto sulla disponibilità sul carico di lavoro.

Le leve di emergenza disabilitano, applicano la limitazione (della larghezza di banda della rete) o modificano il comportamento di componenti o dipendenze mediante meccanismi noti e testati. Ciò può ridurre i danni causati al carico di lavoro dall'esaurimento delle risorse dovuto ad aumenti imprevisti della domanda e l'impatto dei guasti nei componenti non critici all'interno del carico di lavoro.

Risultato desiderato: l'implementazione delle leve di emergenza consente di definire processi noti e validi per mantenere la disponibilità dei componenti critici nel carico di lavoro. Il carico di lavoro dovrebbe diminuire gradualmente e continuare a svolgere le sue funzioni aziendali critiche durante l'attivazione di una leva di emergenza. Per maggiori dettagli sulla degradazione graduale, vedere REL05-BP01 Implementare graceful degradation per trasformare le dipendenze rigide applicabili in dipendenze morbide.

Anti-pattern comuni:

- L'errore a livello di dipendenze non critiche influisce sulla disponibilità del carico di lavoro principale.
- Mancato test o mancata verifica del comportamento dei componenti critici durante il deterioramento delle prestazioni dei componenti non critici.

- Mancata definizione di criteri chiari e deterministici per l'attivazione o la disattivazione di una leva di emergenza.

Vantaggi dell'adozione di questa best practice: l'implementazione delle leve di emergenza migliora la disponibilità dei componenti critici del carico di lavoro fornendo agli addetti alla risoluzione processi consolidati per rispondere a picchi imprevisti della domanda o a guasti delle dipendenze non critiche.

Livello di rischio associato se questa best practice non fosse adottata: medio

Guida all'implementazione

- Identifica i componenti critici del tuo carico di lavoro.
- Progetta e definisci l'architettura dei componenti critici del tuo carico di lavoro in modo che sia in grado di sostenere i guasti dei componenti non critici.
- Esegui i test per convalidare il comportamento dei componenti critici in caso di guasti dei componenti non critici.
- Definisci e monitora le metriche o i trigger pertinenti per avviare le procedure relative alle leve di emergenza.
- Definisci le procedure (manuali o automatiche) che includono la leva di emergenza.

Passaggi dell'implementazione

- Identifica i componenti business-critical nel tuo carico di lavoro.
  - Ogni componente tecnico del carico di lavoro deve essere mappato alla funzione aziendale pertinente e classificato come critico o non critico. Per esempi di funzionalità critiche e non critiche di Amazon, consulta Any Day Can Be Prime Day: How Amazon.com Search Uses Chaos Engineering to Handle Over 84K Requests Per Second.
  - Si tratta di una decisione sia tecnica che aziendale e varia in base all'organizzazione e al carico di lavoro.
- Progetta e definisci l'architettura dei componenti critici del tuo carico di lavoro in modo che sia in grado di sostenere i guasti dei componenti non critici.
  - Durante l'analisi delle dipendenze, valuta tutte le potenziali modalità di guasto e verifica che i meccanismi basati su leve di emergenza forniscano le funzionalità critiche ai componenti a valle.
- Esegui i test per convalidare il comportamento dei componenti critici durante l'attivazione delle leve di emergenza.

  - Evita il comportamento bimodale. Per maggiori dettagli, consulta -BP05 Utilizzare la stabilità statica per prevenire comportamenti bimodali. REL11
- Definisci, monitora e attiva gli avvisi per le metriche pertinenti per avviare la procedura relative alla leva di emergenza.
  - L'individuazione delle metriche da monitorare dipende dal carico di lavoro. Alcuni esempi di metrica sono la latenza o il numero di richieste non riuscite nei confronti di una dipendenza.
- Definisci le procedure (manuali o automatiche) che includono la leva di emergenza.
  - Ciò può includere meccanismi come la riduzione del carico, le richieste di limitazione (della larghezza di banda della rete) o l'implementazione della normale riduzione delle prestazioni.

Risorse

Best practice correlate:

- REL05-BP01 Implementa una degradazione graduale per trasformare le dipendenze rigide applicabili in dipendenze morbide
- REL05-BP02 Richieste Throttle
- REL11-BP05 Usa la stabilità statica per prevenire comportamenti bimodali

Documenti correlati:

- Automatizzazione di distribuzioni pratiche e sicure
- Any Day Can Be Prime Day: How Amazon.com Search Uses Chaos Engineering to Handle Over 84K Requests Per Second

Video correlati:

- AWS re:Invent 2020: affidabilità, coerenza e fiducia grazie all'immutabilità

Gestione delle modifiche

Questions

- REL6. Come si monitorano le risorse del carico di lavoro?
- REL7. Come si progetta il carico di lavoro per adattarsi ai cambiamenti della domanda?

- REL8. In che modo implementare le modifiche?

REL6. Come si monitorano le risorse del carico di lavoro?

I log e le metriche sono strumenti molto efficaci per ottenere informazioni sullo stato del carico di lavoro. Puoi configurare il carico di lavoro in modo da monitorare i log e le metriche e inviare notifiche in caso di superamento delle soglie o di eventi significativi. Il monitoraggio permette al carico di lavoro di riconoscere il superamento delle soglie di prestazioni basse o il verificarsi di errori, in modo da ripristinarlo in automatico di conseguenza.

Best practice

- REL06-BP01 Monitora tutti i componenti per il carico di lavoro (generazione)
- REL06-BP02 Definire e calcolare le metriche (aggregazione)
- REL06-BP03 Invio di notifiche (elaborazione e allarme in tempo reale)
- REL06-BP04 Risposte automatiche (elaborazione e allarme in tempo reale)
- REL06-BP05 Analizza i registri
- REL06-BP06 Effettua revisioni regolarmente
- REL06-BP07 Monitora il end-to-end tracciamento delle richieste attraverso il sistema

REL06-BP01 Monitora tutti i componenti per il carico di lavoro (generazione)

Monitora i componenti del carico di lavoro con Amazon CloudWatch o strumenti di terze parti. Monitora AWS i servizi con AWS Health Dashboard.

Occorre monitorare tutti i componenti del carico di lavoro, inclusi front-end, logica aziendale e livelli di storage. Definisci i parametri chiave e come estrarli dai log, se necessario, e imposta soglie per richiamare gli eventi di allarme corrispondenti. Assicurati che le metriche siano pertinenti agli indicatori chiave di performance (KPIs) del tuo carico di lavoro e utilizza metriche e log per identificare i primi segnali di allarme di degrado del servizio. Ad esempio, una metrica relativa ai risultati aziendali, come il numero di ordini elaborati con successo al minuto, può indicare i problemi relativi al carico di lavoro più rapidamente di una metrica tecnica, come l'utilizzo. CPU Utilizza AWS Health Dashboard per una visualizzazione personalizzata delle prestazioni e della disponibilità dei servizi alla base delle AWS tue risorse. AWS

Il monitoraggio nel cloud offre nuove opportunità. La maggior parte dei provider di servizi cloud ha sviluppato hook personalizzabili e può fornire approfondimenti per aiutarti a monitorare più livelli

del tuo carico di lavoro. AWS servizi come Amazon CloudWatch applicano algoritmi statistici e di apprendimento automatico per analizzare continuamente le metriche di sistemi e applicazioni, determinare linee di base normali e rilevare anomalie con un intervento minimo da parte dell'utente. Gli algoritmi di rilevamento delle anomalie tengono conto delle variazioni di stagionalità e di tendenza dei parametri.

AWS mette a disposizione per il consumo una grande quantità di informazioni di monitoraggio e log che possono essere utilizzate per definire metriche e processi specifici del carico di lavoro e adottare tecniche di apprendimento automatico indipendentemente dall'esperienza nel machine learning. change-in-demand

Inoltre, monitora tutti gli endpoint esterni per avere la certezza che siano indipendenti dall'implementazione di base. Questo monitoraggio attivo può essere svolto attraverso transazioni sintetiche (talvolta definite canary dell'utente, ma da non confondere con le distribuzioni canary) che eseguono periodicamente alcune attività comuni che corrispondono a quelle effettuate dai client del carico di lavoro. Mantieni queste attività di breve durata e assicurati di non sovraccaricare il carico di lavoro durante il test. Amazon CloudWatch Synthetics ti consente di creare canarini sintetici per monitorare i tuoi endpoint e. APIs Puoi anche combinare i nodi client sintetici canary con la console AWS X-Ray per individuare quali canary sintetici stanno riscontrando problemi con errori, guasti o tassi di limitazione (della larghezza di banda della rete) per l'intervallo di tempo selezionato.

Risultato desiderato:

Raccogliere e utilizzare i parametri critici di tutti i componenti del carico di lavoro per garantire l'affidabilità del carico di lavoro e un'esperienza utente ottimale. Rilevare che un carico di lavoro non sta raggiungendo i risultati aziendali consente di dichiarare rapidamente un disastro e di riprendersi da un incidente.

Anti-pattern comuni:

- Solo monitoraggio delle interfacce esterne per il carico di lavoro.
- Non genera parametri specifici per il carico di lavoro e si basa solo sui parametri forniti dai servizi utilizzati dal carico di lavoro. AWS
- Utilizza solo metriche tecniche nel carico di lavoro e non monitora le metriche relative a quelle non tecniche a cui contribuisce il carico di lavoro. KPIs
- Affidarsi al traffico di produzione e a semplici controlli dell'integrità per monitorare e valutare lo stato del carico di lavoro.

Vantaggi dell'adozione di questa best practice: il monitoraggio a tutti i livelli del carico di lavoro consente di prevedere e risolvere più rapidamente i problemi dei componenti che costituiscono il carico di lavoro.

Livello di rischio associato se questa best practice non fosse adottata: elevato

Guida all'implementazione

1. Attiva la creazione di log, laddove possibile. I dati di monitoraggio devono essere ottenuti da tutti i componenti dei carichi di lavoro. Attiva ulteriori log, come i log di accesso S3, e consenti al carico di lavoro di creare log per i dati specifici del carico di lavoro. Raccogli i parametri CPU relativi agli I/O di rete e alle medie di I/O su disco da servizi come Amazon, ECS EKS Amazon, Elastic Load EC2 Balancing e Amazon. AWS Auto Scaling EMR Vedi AWS Servizi che pubblicano CloudWatch metriche per un elenco di servizi su cui pubblicano metriche. AWS CloudWatch
2. Esamina tutti i parametri predefiniti ed esplora eventuali lacune nella raccolta dei dati. Tutti i servizi generano parametri predefiniti. La raccolta di parametri predefiniti consente di comprendere meglio le dipendenze tra i componenti del carico di lavoro e il modo in cui l'affidabilità e le prestazioni dei componenti influiscono sul carico di lavoro. Puoi anche creare e pubblicare le tue metriche CloudWatch utilizzando o un. AWS CLI API
3. Valuta tutte le metriche per decidere quali avvisare per ogni AWS servizio del tuo carico di lavoro. Puoi scegliere di selezionare un sottoinsieme di parametri che hanno un impatto importante sull'affidabilità del carico di lavoro. Concentrarsi su parametri e soglie critiche consente di affinare il numero di avvisi, così da ridurre al minimo i falsi positivi.
4. Definisci gli avvisi e il processo di recupero del carico di lavoro dopo il richiamo dell'avviso. La definizione degli avvisi consente di inviare notifiche, inoltrare e eseguire rapidamente le operazioni necessarie per riprendersi da un incidente e raggiungere l'obiettivo di tempo di ripristino prescritto (). RTO Puoi utilizzare Amazon CloudWatch Alarms per richiamare flussi di lavoro automatizzati e avviare procedure di ripristino basate su soglie definite.
5. Esplora l'uso di transazioni sintetiche per raccogliere dati rilevanti sullo stato dei carichi di lavoro. Il monitoraggio sintetico segue gli stessi percorsi ed esegue le stesse azioni di un cliente, il che consente di verificare continuamente l'esperienza del cliente anche quando non c'è traffico di clienti sui carichi di lavoro. Grazie alle transazioni sintetiche, puoi scoprire i problemi prima che vengano rilevati dai clienti.

Risorse

Best practice correlate:

- REL11-BP03 Automatizza la guarigione su tutti i livelli

Documenti correlati:

- Guida introduttiva alla dashboard: stato del tuo account AWS Health
- AWS Servizi che pubblicano CloudWatch metriche
- Access Logs for Your Network Load Balancer
- Access logs for your application load balancer
- Accesso ad Amazon CloudWatch Logs per AWS Lambda
- Registrazione degli accessi al server Amazon S3
- Enable Access Logs for Your Classic Load Balancer
- Exporting log data to Amazon S3
- Installa l' CloudWatch agente su un'EC2istanza Amazon
- Publishing Custom Metrics
- Utilizzo di Amazon CloudWatch Dashboards
- Utilizzo di Amazon CloudWatch Metrics
- Utilizzo di Canaries (Amazon CloudWatch Synthetics)
- Cosa sono gli Amazon CloudWatch Logs?

Guide per l'utente:

- Creating a trail
- Monitoraggio dei parametri di memoria e disco per le istanze Amazon EC2 Linux
- Utilizzo dei CloudWatch log con istanze di container
- Log di flusso VPC
- Che cos'è Amazon DevOps Guru?
- Che cos'è AWS X-Ray?

Blog correlati:

- Esecuzione del debug con Amazon Synthetics e CloudWatch AWS X-Ray

Esempi e workshop correlati:

- AWS Well-Architected Labs: eccellenza operativa. Monitoraggio delle dipendenze
- The Amazon Builders' Library: strumentazione di sistemi distribuiti per visibilità operativa
- Workshop sull'osservabilità

REL06-BP02 Definire e calcolare le metriche (aggregazione)

Archivia i dati di log e applica i filtri, laddove necessari, per calcolare i parametri, ad esempio i conteggi di un evento del log specifico o la latenza calcolata dai timestamp del log eventi.

Amazon CloudWatch e Amazon S3 fungono da livelli di aggregazione e storage principali. Per alcuni servizi, come AWS Auto Scaling Elastic Load Balancing, per impostazione predefinita vengono fornite metriche predefinite per il CPU carico o la latenza media delle richieste su un cluster o un'istanza. Per i servizi di streaming, come VPC Flow Logs e AWS CloudTrail, i dati degli eventi vengono inoltrati a CloudWatch Logs ed è necessario definire e applicare filtri di metrica per estrarre le metriche dai dati degli eventi. Ciò fornisce dati sulle serie temporali, che possono fungere da input per gli CloudWatch allarmi definiti dall'utente per richiamare gli avvisi.

Livello di rischio associato se questa best practice non fosse adottata: elevato

Guida all'implementazione

- Aggregazione: definisci e calcola i parametri. Archivia i dati di log e applica filtri, se necessario, per calcolare i parametri, ad esempio i conteggi di un evento del log specifico o la latenza calcolata dai timestamp degli eventi di log
  - I filtri metrici definiscono i termini e i modelli da cercare nei dati di registro quando vengono inviati ai registri. CloudWatch CloudWatch Logs utilizza questi filtri metrici per trasformare i dati di registro in CloudWatch metriche numeriche su cui è possibile rappresentare graficamente o impostare un allarme.
    - Ricerca e filtraggio dei dati di log
  - Utilizza una terza parte affidabile per aggregare i log.
    - Segui le istruzioni che ti vengono fornite dalle terze parti. La maggior parte dei prodotti di terze parti si integra con CloudWatch Amazon S3.
  - Alcuni AWS servizi possono pubblicare i log direttamente su Amazon S3. In questo modo, se il tuo requisito principale per i log è l'archiviazione in Amazon S3, potrai facilmente impostare il servizio di produzione dei log affinché li invii direttamente ad Amazon S3 senza configurare un'infrastruttura aggiuntiva.
    - Invio di log direttamente ad Amazon S3

Risorse

Documenti correlati:

- Domande di esempio su Amazon CloudWatch Logs Insights
- Eseguire il debug con Amazon Synthetics e CloudWatch AWS X-Ray
- One Observability Workshop
- Ricerca e filtraggio dei dati di log
- Invio di log direttamente ad Amazon S3
- The Amazon Builders' Library: strumentazione di sistemi distribuiti per visibilità operativa

REL06-BP03 Invio di notifiche (elaborazione e allarme in tempo reale)

Quando le organizzazioni rilevano potenziali problemi, inviano notifiche e avvisi in tempo reale ai team e ai sistemi appropriati per rispondere rapidamente ed efficacemente alle difficoltà.

Risultato desiderato: è possibile rispondere rapidamente agli eventi operativi attraverso la configurazione di allarmi pertinenti in base ai parametri del servizio e dell'applicazione. Quando la soglia degli allarmi viene superata, i team e i sistemi appropriati vengono informati in modo che possano risolvere i problemi sottostanti.

Anti-pattern comuni:

- Configuri gli allarmi con una soglia eccessivamente alta, con conseguente mancato invio di notifiche importanti.
- Configuri gli allarmi con una soglia troppo bassa, con il risultato che gli avvisi importanti non vengono presi in considerazione a causa del numero eccessivo di notifiche generate.
- Non aggiorni gli allarmi e la relativa soglia quando cambia l'utilizzo.
- Per gli allarmi gestiti meglio tramite le azioni automatizzate, l'invio della notifica ai team anziché l'attivazione dell'azione automatizzata comporta la generazione di un numero eccessivo di notifiche.

Vantaggi dell'adozione di questa best practice: l'invio di notifiche e avvisi in tempo reale ai team e ai sistemi appropriati consente di individuare tempestivamente i problemi e di rispondere rapidamente agli incidenti operativi.

Livello di rischio associato se questa best practice non fosse adottata: elevato

Guida all'implementazione

I carichi di lavoro devono essere dotati di sistemi di elaborazione e allarme in tempo reale per migliorare l'identificazione dei problemi che possono influire sulla disponibilità dell'applicazione e fungere da trigger per la risposta automatizzata. Le organizzazioni possono eseguire un sistema di elaborazione e allarme in tempo reale creando avvisi con parametri definiti in modo da ricevere le notifiche ogni volta che si verificano eventi significativi o un parametro supera una determinata soglia.

Amazon ti CloudWatch consente di creare allarmi metrici e compositi utilizzando CloudWatch allarmi basati su soglia statica, rilevamento di anomalie e altri criteri. Per maggiori dettagli sui tipi di allarmi che puoi configurare utilizzando CloudWatch, consulta la sezione sugli allarmi della documentazione. CloudWatch

Puoi creare visualizzazioni personalizzate delle metriche e degli avvisi delle tue AWS risorse per i tuoi team utilizzando i dashboard. CloudWatch Le home page personalizzabili della CloudWatch console consentono di monitorare le risorse in un'unica visualizzazione in più regioni.

Gli allarmi possono eseguire una o più azioni, come inviare una notifica a un SNSargomento di Amazon, eseguire un'EC2azione Amazon o un'azione Amazon EC2 Auto Scaling o creare OpsItem un incidente o. AWS Systems Manager

Amazon CloudWatch utilizza Amazon SNS per inviare notifiche quando lo stato dell'allarme cambia, fornendo la consegna dei messaggi dagli editori (produttori) agli abbonati (consumatori). Per maggiori dettagli sulla configurazione delle SNS notifiche Amazon, consulta Configurazione di Amazon SNS.

CloudWatch invia EventBridgeeventi ogni volta che un CloudWatch allarme viene creato, aggiornato, eliminato o il suo stato cambia. Puoi utilizzare EventBridge questi eventi per creare regole che eseguono azioni, come avvisarti ogni volta che lo stato di un allarme cambia o attivare automaticamente eventi nel tuo account utilizzando l'automazione di Systems Manager.

Quando dovresti usare EventBridge AmazonSNS?

Entrambi EventBridge e Amazon SNS possono essere utilizzati per sviluppare applicazioni basate sugli eventi e la tua scelta dipenderà dalle tue esigenze specifiche.

Amazon EventBridge è consigliato quando desideri creare un'applicazione che reagisca agli eventi delle tue applicazioni, applicazioni SaaS e servizi. AWS EventBridge è l'unico servizio basato su eventi che si integra direttamente con partner SaaS di terze parti. EventBridge inoltre, inserisce automaticamente gli eventi da oltre 200 AWS servizi senza richiedere agli sviluppatori di creare alcuna risorsa nel proprio account.

EventBridge utilizza una struttura JSON basata su una struttura definita per gli eventi e consente di creare regole da applicare all'intero corpo dell'evento per selezionare gli eventi da inoltrare a un obiettivo. EventBridge attualmente supporta oltre 20 AWS servizi come target, tra cui Amazon AWS Lambda, Amazon SQSSNS, Amazon Kinesis Data Streams e Amazon Data Firehose.

Amazon SNS è consigliato per le applicazioni che richiedono un elevato numero di ventole (migliaia o milioni di endpoint). Uno schema comune che vediamo è che i clienti utilizzano Amazon SNS come obiettivo per la loro regola per filtrare gli eventi di cui hanno bisogno e distribuirli a più endpoint.

I messaggi non sono strutturati e possono assumere qualsiasi formato. Amazon SNS supporta l'inoltro di messaggi a sei diversi tipi di destinazioni, tra cui Lambda, SQS Amazon, endpoint SMS /SHTTP, push per dispositivi mobili ed e-mail. La latenza SNS tipica di Amazon è inferiore a 30 millisecondi. Un'ampia gamma di AWS servizi consente di inviare SNS messaggi Amazon configurando il servizio a tale scopo (più di 30, tra cui AmazonEC2, Amazon S3 e Amazon RDS).

Passaggi dell'implementazione

1. Crea un allarme utilizzando Amazon CloudWatch alarms.
    a. Un allarme metrico monitora una singola CloudWatch metrica o un'espressione in base alle metriche. CloudWatch L'allarme avvia una o più azioni in base al valore del parametro o dell'espressione rispetto a una soglia, per un determinato numero di intervalli di tempo. L'azione può consistere nell'invio di una notifica a un SNSargomento di Amazon, nell'esecuzione di un'EC2azione Amazon o di un'azione Amazon EC2 Auto Scaling o nella creazione di un incidente OpsItem o in. AWS Systems Manager
    b. Un allarme composito è costituito da un'espressione di regola che considera le condizioni di altri allarmi che hai creato. L'allarme composito entra in stato di allarme solo se tutte le condizioni della regola sono soddisfatte. Gli allarmi specificati nell'espressione di regola di un allarme composito possono includere allarmi di parametri e allarmi compositi aggiuntivi. Gli allarmi compositi possono inviare SNS notifiche ad Amazon quando il loro stato cambia e possono creare Systems Manager OpsItemso incidenti quando entrano nello stato di allarme, ma non possono eseguire azioni Amazon EC2 o Auto Scaling.
2. Configura SNSle notifiche Amazon. Quando crei un CloudWatch allarme, puoi includere un SNS argomento Amazon per inviare una notifica quando l'allarme cambia stato.
3. Crea regole EventBridge che corrispondano CloudWatch agli allarmi specificati. Ogni regola supporta più destinazioni, incluse le funzioni Lambda. Ad esempio, è possibile definire un allarme che si attiva quando lo spazio disponibile su disco si sta esaurendo, che attiva una

funzione Lambda tramite una EventBridge regola per ripulire lo spazio. Per maggiori dettagli sugli EventBridge obiettivi, vedi obiettivi. EventBridge

Risorse

Best practice Well-Architected correlate:

- REL06-BP01 Monitora tutti i componenti per il carico di lavoro (generazione)
- REL06-BP02 Definire e calcolare le metriche (aggregazione)
- REL12-BP01 Usa i playbook per indagare sui guasti

Documenti correlati:

- Amazon CloudWatch
- CloudWatch Registra approfondimenti
- Utilizzo degli CloudWatch allarmi Amazon
- Utilizzo dei CloudWatch pannelli di controllo di Amazon
- Utilizzo dei CloudWatch parametri di Amazon
- Configurazione SNS delle notifiche Amazon
- CloudWatch rilevamento delle anomalie
- CloudWatch Protezione dei dati dei registri
- Amazon EventBridge
- Amazon Simple Notification Service

Video correlati:

- Video sull'osservabilità reinvent 2022
- AWS re:Invent 2022 - Le migliori pratiche di osservabilità su Amazon

Esempi correlati:

- One Observability Workshop
- Amazon EventBridge punta AWS Lambda al controllo dei feedback tramite Amazon CloudWatch Alarms

REL06-BP04 Risposte automatiche (elaborazione e allarme in tempo reale)

Utilizza l'automazione per agire quando viene rilevato un evento; ad esempio, per sostituire i componenti guasti.

L'elaborazione automatizzata in tempo reale degli allarmi è implementata in modo che i sistemi possano effettuare azioni correttive rapide e tentare di prevenire guasti o danni al servizio quando vengono attivati gli allarmi. Le risposte automatiche agli allarmi potrebbero includere la sostituzione dei componenti guasti, la regolazione della capacità di calcolo, il reindirizzamento del traffico verso host integri, zone di disponibilità o altre regioni e la notifica agli operatori.

Risultato desiderato: vengono identificati gli allarmi in tempo reale e viene impostata l'elaborazione automatica degli allarmi per richiamare le azioni appropriate intraprese per mantenere gli obiettivi dei livelli di servizio e gli accordi sui livelli di servizio (). SLAs L'automazione può interessare un ambito che va dalle attività di autoriparazione dei singoli componenti al failover dell'intero sito.

Anti-pattern comuni:

- Non disporre di un inventario o un catalogo dettagliato dei principali allarmi in tempo reale.
- Nessuna risposta automatica in caso di allarmi critici (ad esempio, quando le risorse di calcolo stanno per esaurirsi, viene implementato il dimensionamento automatico).
- Azioni di risposta agli allarmi contraddittorie.
- Nessuna procedura operativa standard (SOPs) che gli operatori devono seguire quando ricevono notifiche di avviso.
- Non monitorare le modifiche apportate alla configurazione, poiché le modifiche della configurazione non rilevate possono causare tempi di inattività per i carichi di lavoro.
- Non avere una strategia per annullare le modifiche involontarie alla configurazione.

Vantaggi dell'adozione di questa best practice: migliore resilienza del sistema grazie all'automazione dell'elaborazione degli allarmi. Il sistema implementa automaticamente azioni correttive, riducendo le attività manuali che possono comportare interventi umani soggetti a errori. L'operatività del carico di lavoro soddisfa gli obiettivi di disponibilità e riduce le interruzioni del servizio.

Livello di rischio associato se questa best practice non fosse adottata: medio

Guida all'implementazione

Per gestire in modo efficiente gli avvisi e automatizzarne la risposta, classifica gli avvisi in base alla loro criticità e al loro impatto, documenta le procedure di risposta e pianifica le risposte prima di classificare le attività.

Identifica le attività che richiedono azioni specifiche (spesso dettagliate nei runbook) ed esamina tutti i runbook e i playbook per determinare quali attività possono essere automatizzate. Se è possibile definire delle azioni, significa che esse spesso possono essere automatizzate. Se le azioni non possono essere automatizzate, documentate le fasi manuali di un programma SOP e addestrate gli operatori su di esse. Continua ad analizzare dettagliatamente i processi manuali alla ricerca di opportunità di automazione in cui puoi stabilire e mantenere un piano per automatizzare le risposte agli avvisi.

Passaggi dell'implementazione

1. Crea un inventario degli allarmi: per ottenere un elenco di tutti gli allarmi, puoi AWS CLIutilizzare il comando Amazon CloudWatch. `describe-alarms` A seconda del numero di allarmi che hai impostato, potresti dover utilizzare l'impaginazione per recuperare un sottoinsieme di allarmi per ogni chiamata, o in alternativa puoi usare il per AWS SDK ottenere gli allarmi utilizzando una chiamata. API
2. Documenta tutte le azioni associate all'allarme: aggiorna un runbook con tutti gli allarmi e le relative azioni, a prescindere che siano manuali o automatizzati. AWS Systems Manager offre runbook predefiniti. Per informazioni sull'uso dei runbook, consulta Working with runbooks. Per informazioni sulla visualizzazione dei contenuti dei runbook, consulta Visualizza il contenuto del runbook.
3. Configurazione e gestione delle azioni di allarme: per tutti gli allarmi che richiedono un'azione, specifica l'azione automatica utilizzando il. CloudWatch SDK Ad esempio, puoi modificare automaticamente lo stato delle tue EC2 istanze Amazon in base a un CloudWatch allarme creando e abilitando azioni su un allarme o disabilitando le azioni su un allarme.

   Puoi anche utilizzare Amazon EventBridge per rispondere automaticamente a eventi di sistema, come problemi di disponibilità delle applicazioni o modifiche delle risorse. Puoi creare regole che indichino a quali eventi sei interessato e quali operazioni automatizzate eseguire quando un evento corrisponde a una regola. Le azioni che possono essere avviate automaticamente includono l'invocazione di una AWS Lambdafunzione, il richiamo di EC2Run `Command`Amazon, l'inoltro dell'evento ad Amazon Kinesis Data Streams e l'utilizzo di Automate Amazon. EC2 EventBridge

4. Procedure operative standard (SOPs): in base ai componenti dell'applicazione, consiglia più modelli. AWS Resilience HubSOP È possibile utilizzarli SOPs per documentare tutti i processi che un operatore deve seguire nel caso in cui venga generato un avviso. È inoltre possibile crearne uno SOP basato sui consigli di Resilience Hub, in cui è necessaria un'applicazione Resilience Hub con una politica di resilienza associata, nonché una valutazione storica della resilienza rispetto a tale applicazione. I consigli per te SOP sono prodotti dalla valutazione della resilienza.

   Resilience Hub collabora con Systems Manager per automatizzare le fasi del processo SOPs fornendo una serie di SSMdocumenti che è possibile utilizzare come base. SOPs Ad esempio, Resilience Hub può consigliarne uno SOP per aggiungere spazio su disco in base a un documento di automazione esistenteSSM.

5. Esegui azioni automatizzate utilizzando Amazon DevOps Guru: puoi utilizzare Amazon DevOps Guru per monitorare automaticamente le risorse delle applicazioni alla ricerca di comportamenti anomali e fornire consigli mirati per accelerare i tempi di identificazione e risoluzione dei problemi. Con DevOps Guru, puoi monitorare flussi di dati operativi quasi in tempo reale da più fonti, tra cui Amazon CloudWatch Metrics, AWS Config, AWS CloudFormatione. AWS X-Ray Puoi anche usare DevOps Guru per creare OpsItems OpsCenter e inviare automaticamente eventi per EventBridge un'ulteriore automazione.

Risorse

Best practice correlate:

- REL06-BP01 Monitora tutti i componenti per il carico di lavoro (generazione)
- REL06-BP02 Definire e calcolare le metriche (aggregazione)
- REL06-BP03 Invio di notifiche (elaborazione e allarme in tempo reale)
- REL08-BP01 Usa i runbook per attività standard come l'implementazione

Documenti correlati:

- AWS Systems Manager Automazione
- Creazione di una EventBridge regola che si attiva su un evento da una risorsa AWS
- One Observability Workshop
- The Amazon Builders' Library: strumentazione di sistemi distribuiti per visibilità operativa
- Cos'è Amazon DevOps Guru?

- Utilizzo dei documenti di automazione (playbook)

Video correlati:

- AWS re:Invent 2022 - Le migliori pratiche di osservabilità su Amazon
- AWS re:Invent 2020: automatizza qualsiasi cosa con AWS Systems Manager
- Introduzione a AWS Resilience Hub
- Crea sistemi di ticket personalizzati per le notifiche di Amazon DevOps Guru
- Abilita l'aggregazione di informazioni dettagliate su più account con Amazon Guru DevOps

Esempi correlati:

- Workshop sull'affidabilità
- Workshop su Amazon CloudWatch e Systems Manager

REL06-BP05 Analizza i registri

Raccogli i file di log e le cronologie dei parametri e analizzali per ottenere informazioni più ampie sulle tendenze e sui carichi di lavoro.

Amazon CloudWatch Logs Insights supporta un linguaggio di query semplice ma potente che puoi utilizzare per analizzare i dati di log. Amazon CloudWatch Logs supporta anche abbonamenti che consentono ai dati di fluire senza problemi verso Amazon S3, dove puoi utilizzare o Amazon Athena per interrogare i dati. Supporta, inoltre, le query su un'ampia gamma di formati. Per ulteriori informazioni, consulta la sezione Formati supportati SerDes e di dati nella Guida per l'utente di Amazon Athena. Per l'analisi di enormi set di file di log, puoi eseguire un EMR cluster Amazon per eseguire analisi su scala petabyte.

Esistono numerosi strumenti forniti dai AWS partner e da terze parti che consentono l'aggregazione, l'elaborazione, l'archiviazione e l'analisi. Questi strumenti includono New Relic, Splunk, Loggly, Logstash e Nagios. CloudHealth Tuttavia, la generazione esterna di log di sistema e applicazioni è univoca per ciascun provider di servizi cloud e spesso per ciascun servizio.

Una parte spesso trascurata del processo di monitoraggio è la gestione dei dati. È necessario determinare i requisiti di conservazione per il monitoraggio dei dati, quindi applicare le policy del ciclo di vita di conseguenza. Amazon S3 supporta la gestione del ciclo di vita a livello di bucket S3. Questa

gestione del ciclo di vita può essere applicata in modo diverso ai diversi percorsi nel bucket. Verso la fine del ciclo di vita, è possibile trasferire i dati ad Amazon S3 Glacier per lo storage a lungo termine e in seguito la scadenza al termine del periodo di conservazione. La classe di storage S3 Intelligent-Tiering è progettata per ottimizzare i costi trasferendo automaticamente i dati nel livello di accesso più conveniente, senza impatto sulle prestazioni o sovraccarico operativo.

Livello di rischio associato se questa best practice non fosse adottata: medio

Guida all'implementazione

- CloudWatch Logs Insights ti consente di cercare e analizzare in modo interattivo i dati di log in Amazon CloudWatch Logs.
  - Analisi dei dati di log con Logs Insights CloudWatch
  - Domande di esempio su Amazon CloudWatch Logs Insights
- Usa Amazon CloudWatch Logs per inviare i log ad Amazon S3 dove puoi usare o Amazon Athena per interrogare i dati.
  - Come posso usare Amazon Athena per analizzare i log di accesso al server Amazon S3?
    - Crea una policy del ciclo di vita di S3 per il bucket dei log di accesso al server. Configura la policy del ciclo di vita per rimuovere periodicamente i file di log. In questo modo, si riduce la quantità di dati analizzati da Athena per ogni query.
      - Come posso creare una policy del ciclo di vita per un bucket S3?

Risorse

Documenti correlati:

- Domande di esempio su Amazon CloudWatch Logs Insights
- Analisi dei dati di log con Logs Insights CloudWatch
- Esecuzione del debug con Amazon Synthetics e CloudWatch AWS X-Ray
- Come posso creare una policy del ciclo di vita per un bucket S3?
- Come posso usare Amazon Athena per analizzare i log di accesso al server Amazon S3?
- One Observability Workshop
- The Amazon Builders' Library: strumentazione di sistemi distribuiti per visibilità operativa

REL06-BP06 Effettua revisioni regolarmente

Esegui verifiche frequenti delle modalità di implementazione del monitoraggio del carico di lavoro e aggiornalo in base a eventi e modifiche significativi.

Il monitoraggio efficace è basato su parametri aziendali chiave. Assicurati che questi parametri siano presenti nel carico di lavoro man mano che le priorità aziendali cambiano.

L'audit del monitoraggio consente di sapere quando un'applicazione sta raggiungendo gli obiettivi di disponibilità. L'analisi delle cause principali richiede la capacità di scoprire cosa è successo in caso di errori. AWS consente di monitorare lo stato dei tuoi servizi durante un incidente:

- Amazon CloudWatch Logs: puoi archiviare i tuoi log in questo servizio e controllarne il contenuto.
- Amazon CloudWatch Logs Insights: è un servizio completamente gestito che consente di analizzare log di grandi dimensioni in pochi secondi. Offre query e visualizzazioni rapide e interattive.
- AWS Config: permette di vedere quale infrastruttura AWS era in uso in momenti differenti.
- AWS CloudTrail: Puoi vedere quali AWS APIs sono stati richiamati a che ora e da quale principale.

Inoltre AWS, organizziamo una riunione settimanale per esaminare le prestazioni operative e condividere le conoscenze tra i team. Poiché ci sono così tanti team AWS, abbiamo creato The Wheel per scegliere casualmente un carico di lavoro da esaminare. Stabilire una cadenza regolare per le revisioni delle prestazioni operative e la condivisione delle conoscenze migliora la capacità di ottenere prestazioni più elevate dai team operativi.

Anti-pattern comuni:

- Raccolta dei soli parametri predefiniti.
- Impostazione di una strategia di monitoraggio senza alcuna revisione.
- Nessuna discussione sul monitoraggio quando vengono distribuite modifiche importanti.

Vantaggi dell'adozione di questa best practice: la verifica periodica del monitoraggio consente di prevedere potenziali problemi, invece di rispondere alle notifiche quando un problema previsto si verifica effettivamente.

Livello di rischio associato se questa best practice non fosse adottata: medio

Guida all'implementazione

- Crea più pannelli di controllo per il carico di lavoro. Devi disporre di un pannello di controllo di alto livello contenente i parametri aziendali chiave, nonché i parametri tecnici che hai identificato come i più rilevanti per lo stato previsto del carico di lavoro al variare dell'utilizzo. È inoltre importante disporre di pannelli di controllo per vari livelli di applicazione e dipendenze che è possibile ispezionare.
  - Utilizzo di Amazon CloudWatch Dashboards
- Pianifica ed effettua revisioni periodiche dei pannelli di controllo del carico di lavoro. Effettua ispezioni regolari dei pannelli di controllo. La frequenza può essere diversa a seconda di quanto l'ispezione sia approfondita.
  - Ispeziona l'andamento nei parametri. Confronta i valori dei parametri con i valori storici per vedere se ci sono tendenze che potrebbero suggerire l'esame di un particolare aspetto. Riportiamo alcuni esempi: aumento della latenza, riduzione della funzione aziendale primaria e aumento delle risposte all'errore.
  - Identificazione di valori anomali/anomalie nei parametri. Le medie o mediane possono nascondere valori anomali e anomalie. Osserva i valori più alti e più bassi nell'intervallo di tempo e analizza le cause dei risultati estremi. Man mano che continui a eliminare tali cause, la riduzione del numero di valori estremi ti consente di continuare a migliorare la coerenza delle prestazioni del carico di lavoro.
  - Ricerca di bruschi cambiamenti nel comportamento. Un cambiamento repentino della quantità o della direzione di un parametro può indicare un cambiamento nell'applicazione o fattori esterni che potrebbero richiedere l'aggiunta di ulteriori parametri da monitorare.

Risorse

Documenti correlati:

- Domande di esempio su Amazon CloudWatch Logs Insights
- Esecuzione del debug con Amazon Synthetics e CloudWatch AWS X-Ray
- One Observability Workshop
- The Amazon Builders' Library: strumentazione di sistemi distribuiti per visibilità operativa
- Utilizzo di Amazon CloudWatch Dashboards

REL06-BP07 Monitora il end-to-end tracciamento delle richieste attraverso il sistema

Tieni traccia delle richieste durante l'elaborazione dei componenti del servizio in modo che i team del prodotto possano analizzare i problemi, semplificarne il debug e migliorare le prestazioni.

Risultato desiderato: i carichi di lavoro con tracciamento completo su tutti i componenti sono facili da eseguire il debug, migliorando il tempo medio di risoluzione (MTTR) degli errori e la latenza semplificando l'individuazione della causa principale. End-to-endil tracciamento riduce il tempo necessario per scoprire i componenti interessati e approfondire in dettaglio le cause profonde degli errori o della latenza.

Anti-pattern comuni:

- Il tracciamento viene utilizzato per alcuni componenti ma non per tutti. Ad esempio, senza tracciamento, i team potrebbero non comprendere chiaramente la AWS Lambda latenza causata dagli avviamenti a freddo in un carico di lavoro con picchi di lavoro.
- I canarini sintetici o il monitoraggio degli utenti reali () RUM non sono configurati con il tracciamento. Senza canaries orRUM, la telemetria relativa all'interazione con il cliente viene omessa dall'analisi delle tracce, producendo un profilo prestazionale incompleto.
- I carichi di lavoro ibridi includono strumenti di tracciamento cloud-native (nativi del cloud) e di terze parti, ma non sono state prese misure specifiche per selezionare e integrare completamente un'unica soluzione di tracciamento. In base alla soluzione di tracciamento scelta, è necessario utilizzare il tracciamento nativo del cloud per strumentare componenti che non sono nativi del cloud oppure gli strumenti di terze parti SDKs devono essere configurati per assimilare la telemetria di tracciamento nativa del cloud.

Vantaggi dell'adozione di questa best practice: quando vengono avvisati della presenza di problemi, i team di sviluppo possono visualizzare un quadro completo delle interazioni tra i componenti del sistema, inclusa la correlazione componente per componente con creazione di log, prestazioni e guasti. Poiché il tracciamento semplifica l'identificazione visiva delle cause principali, viene dedicato meno tempo all'individuazione di tali cause. I team che hanno una visione dettagliata delle interazioni tra i componenti prendono decisioni migliori e più rapide durante la fase di risoluzione dei problemi. Le decisioni, ad esempio quando richiamare il failover del ripristino di emergenza o dove implementare in modo più efficace le strategie di riparazione automatica, possono essere migliorate analizzando le tracce dei sistemi; ciò ottimizza in ultima analisi la soddisfazione dei clienti nei confronti dei servizi.

Livello di rischio associato se questa best practice non fosse adottata: medio

Guida all'implementazione

I team che gestiscono le applicazioni distribuite possono utilizzare strumenti di tracciamento per definire un identificatore di correlazione, raccogliere le tracce delle richieste e creare mappe di servizio dei componenti connessi. Tutti i componenti dell'applicazione devono essere inclusi nelle tracce delle richieste, inclusi client di servizio, gateway middleware (software intermediario) e router di eventi, componenti di elaborazione e archiviazione, tra cui gli archivi e i database dei valori chiave. Includi canarini sintetici e il monitoraggio di utenti reali nella configurazione di end-to-end tracciamento per misurare le interazioni e la latenza dei clienti remoti in modo da poter valutare con precisione le prestazioni dei sistemi rispetto agli accordi e agli obiettivi sui livelli di servizio.

Puoi utilizzare AWS X-Rayi servizi di strumentazione Amazon CloudWatch Application Monitoring per fornire una visione completa delle richieste mentre viaggiano attraverso la tua applicazione. X-Ray raccoglie la telemetria delle applicazioni e consente di visualizzarla e filtrarla tra payload, funzioni, tracceAPIs, servizi e può essere attivata per componenti di sistema senza codice o low-code. CloudWatch il monitoraggio delle applicazioni include l'integrazione delle tracce con ServiceLens metriche, registri e allarmi. CloudWatch il monitoraggio delle applicazioni include anche strumenti sintetici per monitorare gli endpoint eAPIs, oltre al monitoraggio degli utenti reali, per strumentare i client delle applicazioni Web.

Passaggi dell'implementazione

- AWS X-Ray Utilizzalo su tutti i servizi nativi supportati come Amazon S3 e Amazon AWS Lambda API Gateway. Questi AWS servizi abilitano X-Ray con toggle di configurazione utilizzando l'infrastruttura come codice, AWS SDKs oppure. AWS Management Console
- Dota di strumenti le applicazioni AWS Distro for Open Telemetry e X-Ray o gli agenti di raccolta di terze parti.
- Consulta la Guida per gli sviluppatori AWS X-Ray per l'implementazione specifica del linguaggio di programmazione. Queste sezioni della documentazione descrivono in dettaglio come strumentare HTTP richieste, SQL interrogazioni e altri processi specifici del linguaggio di programmazione delle applicazioni.
- Usa il tracciamento X-Ray per Amazon CloudWatch Synthetic Canaries e CloudWatch RUMAmazon per analizzare il percorso della richiesta dal tuo client utente finale all'infrastruttura downstream. AWS
- Configura CloudWatch metriche e allarmi in base allo stato delle risorse e alla telemetria di Canary in modo che i team vengano avvisati rapidamente dei problemi e possano quindi approfondire le tracce e le mappe dei servizi con. ServiceLens

- Abilita l'integrazione X-Ray per gli strumenti di tracciamento di terze parti come Datadog, New Relic o Dynatrace in caso di utilizzo di strumenti di terze parti per la tua soluzione di tracciamento principale.

Risorse

Best practice correlate:

- REL06-BP01 Monitora tutti i componenti per il carico di lavoro (generazione)
- REL11-BP01 Monitora tutti i componenti del carico di lavoro per rilevare i guasti

Documenti correlati:

- Che cos'è? AWS X-Ray
- Amazon CloudWatch: monitoraggio delle applicazioni
- Esecuzione del debug con Amazon Synthetics e CloudWatch AWS X-Ray
- The Amazon Builders' Library: strumentazione di sistemi distribuiti per visibilità operativa
- Integrazione AWS X-RayAWS con altri servizi
- AWS Distro per e OpenTelemetry AWS X-Ray
- Amazon CloudWatch: utilizzo del monitoraggio sintetico
- Amazon CloudWatch: utilizzo CloudWatch RUM
- Configura Amazon CloudWatch synthetics canary e Amazon Alarm CloudWatch
- Disponibilità e oltre: comprensione e miglioramento della resilienza dei sistemi distribuiti su AWS

Esempi correlati:

- One Observability Workshop

Video correlati:

- AWS re:Invent 2022 - Come monitorare le applicazioni su più account
- Come monitorare le tue applicazioni AWS

Strumenti correlati:

- AWS X-Ray
- Amazon CloudWatch
- Amazon Route 53

REL7. Come si progetta il carico di lavoro per adattarsi ai cambiamenti della domanda?

Un carico di lavoro scalabile garantisce l'elasticità per aggiungere o rimuovere risorse in automatico, in modo che sussista una stretta corrispondenza con la domanda attuale in un dato momento.

Best practice

- REL07-BP01 Usa l'automazione per ottenere o scalare le risorse
- REL07-BP02 Ottenere risorse dopo aver rilevato una compromissione del carico di lavoro
- REL07-BP03 Ottieni risorse dopo aver rilevato che sono necessarie più risorse per un carico di lavoro
- REL07-BP04 Load Testa il tuo carico di lavoro

REL07-BP01 Usa l'automazione per ottenere o scalare le risorse

Quando sostituisci risorse compromesse o ridimensioni il carico di lavoro, automatizza il processo utilizzando AWS servizi gestiti, come Amazon S3 e. AWS Auto Scaling Puoi anche utilizzare strumenti di terze parti e automatizzare il ridimensionamento. AWS SDKs

AWS I servizi gestiti includono Amazon S3 CloudFront, Amazon, AWS Auto Scaling AWS Lambda, Amazon DynamoDB e Amazon Route AWS Fargate 53.

AWS Auto Scaling consente di rilevare e sostituire le istanze danneggiate. Consente inoltre di creare piani di scalabilità per risorse tra cui EC2 istanze Amazon e flotte Spot, ECS attività Amazon, tabelle e indici Amazon DynamoDB e repliche Amazon Aurora.

Quando scalate le EC2 istanze, assicuratevi di utilizzare più zone di disponibilità (preferibilmente almeno tre) e aggiungete o rimuovete capacità per mantenere l'equilibrio tra queste zone di disponibilità. ECSIe attività o i pod Kubernetes (quando si utilizza Amazon Elastic Kubernetes Service) devono inoltre essere distribuiti su più zone di disponibilità.

Quando vengono utilizzate, le istanze si ridimensionano automaticamente. AWS Lambda Ogni volta che viene ricevuta una notifica di evento relativa alla funzione, individua AWS Lambda rapidamente

la capacità disponibile all'interno del parco di elaborazione ed esegue il codice fino alla concorrenza assegnata. Devi assicurarti che la simultaneità necessaria sia configurata sulla Lambda specifica e in Service Quotas.

Amazon S3 procede a scalare in automatico le risorse per gestire elevati tassi di richiesta. Ad esempio, l'applicazione può soddisfare almeno 3.500///DELETEo 5.500 PUT COPY POSTGET/HEADrichieste al secondo per prefisso in un bucket. Non esistono limiti al numero di prefissi in un bucket. È possibile aumentare le proprie performance in lettura o scrittura parallelizzando le scritture. Ad esempio, se si creano 10 prefissi in un bucket Amazon S3 per parallelizzare le letture, è possibile scalare le prestazioni di lettura a 55.000 richieste di lettura al secondo.

Configura e usa Amazon CloudFront o una rete di distribuzione di contenuti affidabile (CDN). A CDN può fornire tempi di risposta più rapidi per gli utenti finali e soddisfare le richieste di contenuti dalla cache, riducendo così la necessità di scalare il carico di lavoro.

Anti-pattern comuni:

- Implementare gruppi Auto Scaling per la correzione automatica, ma senza elasticità.
- Utilizzare il dimensionamento automatico per rispondere a grandi aumenti di traffico.
- Implementare applicazioni altamente stateful, eliminando l'opzione di elasticità.

Vantaggi dell'adozione di questa best practice: l'automazione elimina il potenziale di errori manuali nell'implementazione e nella disattivazione delle risorse. L'automazione elimina il rischio di superamento dei costi e di rifiuto del servizio a causa della risposta lenta alle esigenze di implementazione o disattivazione.

Livello di rischio associato se questa best practice non fosse adottata: elevato

Guida all'implementazione

- Configura e utilizza AWS Auto Scaling. In questo modo è possibile monitorare le applicazioni e regolare automaticamente la capacità per mantenere prestazioni stabili e prevedibili al minor costo possibile. Grazie ad AWS Auto Scaling, puoi configurare il dimensionamento delle applicazioni per più risorse in vari servizi.
  - Che cos'è AWS Auto Scaling?
    - Configura Auto Scaling su EC2 istanze Amazon e flotte Spot, attività Amazon, tabelle e indici ECS Amazon DynamoDB, repliche Amazon Aurora e appliance, a seconda dei casi. Marketplace AWS

    - Managing throughput capacity automatically with DynamoDB Auto Scaling
        - Utilizza API le operazioni di servizio per specificare gli allarmi, le politiche di scalabilità, i tempi di riscaldamento e i tempi di raffreddamento.
- Utilizza Elastic Load Balancing I bilanciatori del carico possono distribuire il carico in base al percorso o alla connettività di rete.
    - Cos'è l'Elastic Load Balancing?
        - Un Application Load Balancer può distribuire il carico in base al percorso.
            - Cos'è un Application Load Balancer?
                - Configura un Application Load Balancer per distribuire il traffico su diversi carichi di lavoro in base a un percorso nello stesso nome di dominio.
                - Gli Application Load Balancer possono essere utilizzati per distribuire i carichi in modo da integrarsi con la gestione della domanda. AWS Auto Scaling
                    - Uso di un bilanciatore del carico con un gruppo Auto Scaling
        - I Network Load Balancer possono distribuire il carico in base alla connessione.
            - Cos'è un Network Load Balancer?
                - Configura un Network Load Balancer per distribuire il traffico su diversi carichi di lavoro utilizzando TCP o per avere un set costante di indirizzi IP per il tuo carico di lavoro.
                - I Network Load Balancer possono essere utilizzati per distribuire i carichi in modo da integrarsi con la gestione della domanda. AWS Auto Scaling
- Utilizza un provider ad alta disponibilità. DNS DNSi nomi consentono agli utenti di inserire nomi anziché indirizzi IP per accedere ai carichi di lavoro e distribuiscono queste informazioni in un ambito definito, di solito a livello globale per gli utenti del carico di lavoro.
    - Usa Amazon Route 53 o un DNS provider affidabile.
        - What is Amazon Route 53?
    - Usa Route 53 per gestire le CloudFront distribuzioni e i sistemi di bilanciamento del carico.
        - Individua i domini e i sottodomini da gestire.
        - Crea set di record appropriati utilizzando ALIAS i nostri record. CNAME
            - Utilizzo dei record
- Utilizzate la rete AWS globale per ottimizzare il percorso dagli utenti alle applicazioni. AWS Global Accelerator monitora continuamente lo stato degli endpoint delle applicazioni e reindirizza il traffico verso endpoint integri in meno di 30 secondi.

    - AWS Global Accelerator è un servizio che migliora la disponibilità e le prestazioni delle applicazioni con utenti locali o globali. Fornisce indirizzi IP statici che fungono da punto di ingresso fisso agli endpoint delle applicazioni in uno o più dispositivi Regioni AWS, come Application Load Balancer, Network Load Balancer o istanze Amazon. EC2
        - What Is AWS Global Accelerator?
- Configura e usa Amazon CloudFront o una rete di distribuzione di contenuti affidabile (CDN). Una rete di distribuzione di contenuti (CDN) può fornire tempi di risposta più rapidi agli utenti finali e soddisfare richieste di contenuti che possono causare un dimensionamento non necessario dei carichi di lavoro.
    - Che cos'è Amazon CloudFront?
        - Configura CloudFront le distribuzioni Amazon per i tuoi carichi di lavoro o utilizza una terza parte. CDN
            - Puoi limitare l'accesso ai tuoi carichi di lavoro in modo che siano accessibili solo CloudFront utilizzando gli intervalli di IP utilizzati nei gruppi di sicurezza degli endpoint o CloudFront nelle policy di accesso.

Risorse

Documenti correlati:

- APNPartner: partner che possono aiutarti a creare soluzioni di elaborazione automatizzate
- AWS Auto Scaling: How Scaling Plans Work
- Marketplace AWS: prodotti che possono essere utilizzati con Auto Scaling
- Managing Throughput Capacity Automatically with DynamoDB Auto Scaling
- Uso di un bilanciatore del carico con un gruppo Auto Scaling
- Che cos'è AWS Global Accelerator?
- Che cos'è Amazon EC2 Auto Scaling?
- Che cos'è AWS Auto Scaling?
- Che cos'è Amazon CloudFront?
- What is Amazon Route 53?
- Cos'è l'Elastic Load Balancing?
- Cos'è un Network Load Balancer?
- Cos'è un Application Load Balancer?

- Utilizzo dei record

REL07-BP02 Ottenere risorse dopo aver rilevato una compromissione del carico di lavoro

All'occorrenza, procedi a scalare le risorse in modo reattivo se la disponibilità è influenzata per ripristinare la disponibilità del carico di lavoro.

Devi prima configurare il controllo dell'integrità e i criteri su questi controlli per indicare quando la disponibilità è influenzata dalla mancanza di risorse. Quindi invita il personale appropriato a scalare manualmente la risorsa o attivare l'automazione per dimensionarla automaticamente.

La scalabilità può essere regolata manualmente in base al carico di lavoro (ad esempio, modificando il numero di EC2 istanze in un gruppo Auto Scaling o modificando il throughput di una tabella DynamoDB tramite o). AWS Management Console AWS CLI Tuttavia, è opportuno ricorrere all'automazione ogni volta che è possibile (consulta Utilizzo dell'automazione per l'acquisizione o il dimensionamento delle risorse).

Risultato desiderato: avvio di operazioni di dimensionamento (in automatico o manualmente) per il ripristino della disponibilità in caso di rilevamento di un guasto o di un peggioramento dell'esperienza del cliente.

Livello di rischio associato se questa best practice non fosse adottata: medio

Guida all'implementazione

Implementa l'osservabilità e il monitoraggio su tutti i componenti del carico di lavoro, per monitorare l'esperienza del cliente e rilevare i guasti. Definisci le procedure, manuali o automatizzate, che ridimensionano le risorse richieste. o Per ulteriori informazioni, consulta REL11-BP01 Monitora tutti i componenti del carico di lavoro per rilevare eventuali guasti.

Passaggi dell'implementazione

- Definisci le procedure (manuali o automatiche) per scalare le risorse richieste.
  - Le procedure di dimensionamento dipendono da come sono progettati i diversi componenti del carico di lavoro.
  - Le procedure di dimensionamento variano anche a seconda della tecnologia sottostante utilizzata.
    - I componenti utilizzati AWS Auto Scaling possono utilizzare piani di scalabilità per configurare una serie di istruzioni per scalare le risorse. Se utilizzi AWS CloudFormation o aggiungi tag alle AWS risorse, puoi impostare piani di ridimensionamento per diversi set di risorse

per applicazione. Auto Scaling fornisce raccomandazioni per strategie di dimensionamento personalizzate per ogni risorsa. Dopo aver creato il piano di dimensionamento, Auto Scaling combina i metodi di dimensionamento dinamico e predittivo per supportare la tua strategia di dimensionamento. Per ulteriori informazioni, consulta How scaling plans work.

- Amazon EC2 Auto Scaling verifica che tu abbia il numero corretto di EC2 istanze Amazon disponibili per gestire il carico della tua applicazione. Si creano raccolte di EC2 istanze, chiamate gruppi di Auto Scaling. Puoi specificare il numero minimo e massimo di istanze in ogni gruppo di Auto Scaling e Amazon Auto EC2 Scaling garantisce che il tuo gruppo non superi o superi mai questi limiti. Per ulteriori dettagli, consulta Cos'è Amazon EC2 Auto Scaling?
- La scalabilità automatica di Amazon DynamoDB utilizza il servizio Application Auto Scaling per regolare in modo dinamico la capacità effettiva di trasmissione allocata per conto tuo in risposta ai modelli di traffico effettivi. In tal modo una tabella o un indice secondario globale può aumentare la capacità di lettura e scrittura allocata per gestire improvvisi aumenti di traffico, senza limitazione (della larghezza di banda della rete). Per ulteriori dettagli, consulta Managing throughput capacity automatically with DynamoDB auto scaling.

Risorse

Best practice correlate:

- REL07-BP01 Usa l'automazione per ottenere o scalare le risorse
- REL11-BP01 Monitora tutti i componenti del carico di lavoro per rilevare i guasti

Documenti correlati:

- AWS Auto Scaling: How Scaling Plans Work
- Managing Throughput Capacity Automatically with DynamoDB Auto Scaling
- Che cos'è Amazon EC2 Auto Scaling?

REL07-BP03 Ottieni risorse dopo aver rilevato che sono necessarie più risorse per un carico di lavoro

Dimensiona le risorse in modo proattivo per soddisfare la domanda ed evitare l'impatto sulla disponibilità.

Molti AWS servizi si scalano automaticamente per soddisfare la domanda. Se utilizzi EC2 istanze Amazon o ECS cluster Amazon, puoi configurare il ridimensionamento automatico di questi in

modo che avvenga in base a parametri di utilizzo che corrispondono alla domanda per il tuo carico di lavoro. Per AmazonEC2, è possibile CPU utilizzare l'utilizzo medio, il numero di richieste di bilanciamento del carico o la larghezza di banda di rete per scalare orizzontalmente (o scalare in) istanze. EC2 Per AmazonECS, CPU l'utilizzo medio, il numero di richieste di bilanciamento del carico e l'utilizzo della memoria possono essere utilizzati per attività con scalabilità orizzontale (o scalabile). ECS Se si attiva Target Auto Scaling AWS, la scala automatica agisce come un termostato domestico, aggiungendo o rimuovendo risorse per mantenere il valore target (ad esempio, il 70% di utilizzo) specificato. CPU

Amazon EC2 Auto Scaling può anche eseguire Predictive Auto Scaling, che utilizza l'apprendimento automatico per analizzare il carico di lavoro storico di ogni risorsa e prevede regolarmente il carico futuro.

La legge di Little aiuta a calcolare quante istanze di calcolo (EC2istanze, funzioni Lambda simultanee, ecc.) sono necessarie.

$L = \lambda W$

L = numero di istanze (o simultaneità media nel sistema)

λ = velocità media alla quale arrivano le richieste (richieste/sec)

W = tempo medio trascorso da ogni richiesta nel sistema (sec)

Ad esempio, a 100 rps, se ogni richiesta impiega 0,5 secondi per l'elaborazione, avrai bisogno di 50 istanze per tenere il passo con la domanda.

Livello di rischio associato se questa best practice non fosse adottata: medio

Guida all'implementazione

- Ottieni risorse dopo aver rilevato che sono necessarie più risorse per un carico di lavoro Dimensiona le risorse in modo proattivo per soddisfare la domanda ed evitare l'impatto sulla disponibilità.
  - Valuta quante risorse di calcolo sono necessarie (simultaneità di calcolo) per gestire un determinato tasso di richiesta
    - Telling Stories About Little's Law
  - Se disponi di un modello storico di utilizzo, configura la scalabilità pianificata per Amazon EC2 auto scaling.

    - Scalabilità pianificata per Amazon EC2 Auto Scaling
- Usa la scalabilità AWS predittiva.
    - Scalabilità predittiva per Amazon EC2 Auto Scaling

Risorse

Documenti correlati:

- Marketplace AWS: prodotti che possono essere utilizzati con Auto Scaling
- Managing Throughput Capacity Automatically with DynamoDB Auto Scaling
- Scalabilità predittiva perEC2, basata sul Machine Learning
- Scalabilità pianificata per Amazon EC2 Auto Scaling
- Telling Stories About Little's Law
- Che cos'è Amazon EC2 Auto Scaling?

REL07-BP04 Load Testa il tuo carico di lavoro

Adotta un metodo di test del carico per misurare se l'attività di dimensionamento soddisfa i requisiti del carico di lavoro.

È importante eseguire test di carico prolungati. I test di carico dovrebbero scoprire il punto di rottura e testare le prestazioni del carico di lavoro. AWS semplifica la configurazione di ambienti di test temporanei che modellano la scala del carico di lavoro di produzione. Nel cloud, puoi creare un ambiente di test su scala produttiva on demand, completare i test e disattivare le risorse. Poiché paghi per l'ambiente di test solo quando è in esecuzione, puoi simulare un ambiente live a un costo notevolmente inferiore rispetti ai test on-premises.

I test di carico in produzione dovrebbero anche essere considerati come parte delle giornate di gioco in cui il sistema di produzione viene messo alla prova, durante le ore di utilizzo inferiore del cliente, con tutto il personale a disposizione per interpretare i risultati e risolvere eventuali problemi che si presentano.

Anti-pattern comuni:

- Eseguire test di carico su implementazioni che non presentano la stessa configurazione della tua produzione.

- Eseguire test di carico solo su singole parti del carico di lavoro e non sulla sua interezza.
- Eseguire test di carico con un sottoinsieme di richieste e non con un set rappresentativo delle richieste effettive.
- Eseguire test di carico su un fattore di sicurezza di poco superiore al carico previsto.

Vantaggi dell'adozione di questa best practice: saprai quali sono i componenti dell'architettura che non funzionano sotto carico e potrai identificare per tempo i parametri che indicano l'avvicinamento al carico in questione, così da affrontare il problema e prevenire l'impatto dell'esito negativo.

Livello di rischio associato se questa best practice non fosse adottata: medio

Guida all'implementazione

- Esegui test di carico per identificare quali aspetti del carico di lavoro indicano la necessità di aggiungere o rimuovere capacità. Il test di carico deve avere un traffico rappresentativo simile a quello che ricevi nella produzione. Aumenta il carico mentre osservi i parametri implementati per stabilire quale di questi indica quando è necessario aggiungere o rimuovere risorse.
  - Test di carico distribuito su AWS: simula migliaia di utenti connessi
    - Identifica la combinazione di richieste. Potresti avere diverse combinazioni di richieste, quindi dovresti esaminare vari intervalli di tempo per identificare la combinazione di traffico.
    - Implementa un driver di caricamento. Puoi utilizzare codice personalizzato, software open source o software commerciale per implementare un driver di carico.
    - Esegui un test di carico iniziale con una capacità ridotta. Puoi vedere alcuni effetti immediati applicando il carico su una capacità inferiore, possibilmente pari a un'istanza o a un container.
    - Esegui un test di carico con una capacità maggiore. Gli effetti saranno diversi su un carico distribuito, quindi è necessario eseguire il test in condizioni quanto più simili possibili all'ambiente del prodotto.

Risorse

Documenti correlati:

- Test di carico distribuito su AWS: simula migliaia di utenti connessi
- Load testing applications

Video correlati:

- AWS Summit ANZ 2023: accelera con sicurezza grazie ai test di carico AWS distribuiti

REL8. In che modo implementare le modifiche?

Per implementare nuove funzionalità e verificare che i carichi di lavoro e l'ambiente operativo eseguano software noti e che sia possibile applicare patch o sostituirli in modo prevedibile, sono necessarie modifiche controllate. Se invece non sono controllate, risulta difficile prevederne l'effetto o risolvere eventuali problemi che causano.

Best practice

- REL08-BP01 Usa i runbook per attività standard come l'implementazione
- REL08-BP02 Integra i test funzionali come parte della tua implementazione
- REL08-BP03 Integra i test di resilienza come parte della tua implementazione
- REL08-BP04 Implementazione utilizzando un'infrastruttura immutabile
- REL08-BP05 Implementa le modifiche con l'automazione

REL08-BP01 Usa i runbook per attività standard come l'implementazione

I runbook sono le procedure predefinite per ottenere risultati specifici. Utilizza i runbook per eseguire attività standard, o manualmente o automaticamente. Gli esempi includono la distribuzione di un carico di lavoro, l'applicazione di patch a un carico di lavoro o l'esecuzione di modifiche. DNS

Ad esempio, mettere in atto processi per garantire la sicurezza del rollback durante le implementazioni. Garantire la possibilità di eseguire il rollback di un'implementazione senza interruzioni per i clienti è fondamentale per rendere un servizio affidabile.

Per le procedure di runbook, inizia da un processo manuale valido ed efficace, implementalo nel codice e richiamalo per l'esecuzione automatica, se necessario.

Anche per carichi di lavoro sofisticati e altamente automatizzati, i runbook sono ancora utili per l'esecuzione di giornate di gioco o per soddisfare rigorosi requisiti di reportistica e audit.

Tieni presente che i playbook vengono utilizzati in risposta a incidenti specifici e i runbook vengono utilizzati per ottenere risultati specifici. Spesso, i runbook sono per attività di routine, mentre i playbook vengono utilizzati per rispondere a eventi non di routine.

Anti-pattern comuni:

- Eseguire modifiche impreviste alla configurazione nella produzione.
- Ignorare le fasi del piano per velocizzare l'implementazione, compromettendone la riuscita.
- Apportare modifiche senza testarne l'annullamento.

Vantaggi dell'adozione di questa best practice: la pianificazione efficace aumenta la capacità di eseguire correttamente le modifiche, perché sei a conoscenza di tutti i sistemi interessati. Convalidare le modifiche negli ambienti di test aumenta la tua sicurezza.

Livello di rischio associato se questa best practice non fosse adottata: elevato

Guida all'implementazione

- Fornisci risposte coerenti e tempestive a eventi noti documentando le procedure nei runbook.
  - Framework AWS Well-Architected, concetti: runbook
- Usa il principio di Infrastructure as code per definire l'infrastruttura Utilizzando AWS CloudFormation (o una terza parte affidabile) per definire l'infrastruttura, è possibile utilizzare il software di controllo della versione per creare versioni e tenere traccia delle modifiche.
  - Utilizza AWS CloudFormation (o un provider terzo affidabile) per definire la tua infrastruttura.
    - Che cos'è AWS CloudFormation?
  - Crea modelli unici e disaccoppiati, utilizzando solidi principi di progettazione del software.
    - Stabilisci le autorizzazioni, i modelli e le parti responsabili dell'implementazione
      - Controllo degli accessi con AWS Identity and Access Management
    - Usa il controllo del codice sorgente, ad esempio AWS CodeCommit uno strumento affidabile di terze parti, per il controllo delle versioni.
      - Che cos'è AWS CodeCommit?

Risorse

Documenti correlati:

- APNPartner: partner che possono aiutarti a creare soluzioni di implementazione automatizzate
- Marketplace AWS: prodotti per l'automazione delle implementazioni
- AWS Well-Architected Framework: Concetti: Runbook
- Che cos'è? AWS CloudFormation

- Che cos'è AWS CodeCommit?

Esempi correlati:

- Automazione delle operazioni con playbook e runbook

REL08-BP02 Integra i test funzionali come parte della tua implementazione

L'esecuzione di test funzionali è parte integrante dell'implementazione automatizzata. Se non si soddisfano i criteri di esito positivo, la pipeline si arresta o viene sottoposta a rollback. Questi test vengono eseguiti in un ambiente di pre-produzione, gestito per fasi prima della produzione nella pipeline. Idealmente, questa operazione viene eseguita come parte di una pipeline di implementazione.

Risultato desiderato: utilizzi l'automazione per eseguire test funzionali e i dati dei test associati riducono la durata e il costo dei test e migliorano l'accuratezza dei risultati. Integri i test funzionali come parte del processo di implementazione per automatizzare le pipeline di rilascio per l'esecuzione di aggiornamenti rapidi e affidabili di applicazioni e infrastrutture.

Anti-pattern comuni:

- I test vengono eseguiti manualmente al di fuori della pipeline di implementazione.
- Non esegui le fasi di test nell'automazione tramite flussi di lavoro manuali di emergenza.
- Non segui i piani e i processi di test stabiliti a favore di tempistiche accelerate.

Vantaggi dell'adozione di questa best practice: convalida del funzionamento del sistema in base a requisiti specifici grazie ai test funzionali. Viene utilizzato per verificare costantemente l'ordine di funzionamento previsto di componenti come interfacce utenteAPIs, database e codice sorgente. Quando esamini questi componenti del sistema, i test funzionali verificano che ciascuna funzionalità si comporti come previsto, tutelando sia le esigenze degli utenti sia l'integrità del software. Integra i test funzionali come parte dell'implementazione regolare e utilizza l'automazione per implementare tutte le modifiche, riducendo i potenziali errori umani.

Livello di rischio associato se questa best practice non fosse adottata: elevato

Guida all'implementazione

Esegui test funzionali come parte integrante dell'implementazione. L'esecuzione di test funzionali è parte integrante dell'implementazione automatizzata. Se i criteri di successo non vengono soddisfatti,

la pipeline viene interrotta o ripristinata. AWS CodePipeline fornisce una pipeline di distribuzione continua per i test automatizzati, che consente ai tester di automatizzare l'intero processo di test e implementazione. Si integra con AWS servizi come AWS CodeBuild e AWS CodeDeploy per automatizzare le fasi di creazione, test e implementazione del ciclo di vita dello sviluppo del software.

Passaggi dell'implementazione

- Configura la tua pipeline: configura le fasi di origine, compilazione, test e distribuzione utilizzando la console o (). AWS CodePipeline AWS Command Line Interface CLI
  - Definisci la tua fonte: con AWS CodePipeline, puoi recuperare automaticamente il codice sorgente da sistemi di controllo delle versioni come GitHub, o Bitbucket AWS CodeCommit, che verifica che per i test venga sempre utilizzato il codice più recente.
  - Automatizza build e test: AWS CodeBuild puoi creare e testare automaticamente il tuo codice e generare report di test. Supporta i framework di test più diffusi come JUnitNUnit, e TestNG.
  - Distribuisci il tuo codice: una volta che il codice è stato creato e testato, AWS CodeDeploy puoi distribuirlo nel tuo ambiente di test, tra cui EC2 istanze, AWS Lambda funzioni o server locali di Amazon.
  - Monitora le pipeline: AWS CodePipeline ti permette di tenere traccia dei progressi della tua pipeline e dello stato di ciascuna fase. Puoi anche utilizzare i controlli di qualità per bloccare la pipeline in base allo stato di esecuzione dei test. Puoi inoltre ricevere notifiche per qualsiasi errore che si verifica durante l'esecuzione o il completamento della pipeline.

Risorse

Documenti correlati:

- Usalo AWS CodePipeline con AWS CodeBuild per testare il codice ed eseguire build
- Registrazione e monitoraggio AWS CodeBuild
- Indicators for functional testing

REL08-BP03 Integra i test di resilienza come parte della tua implementazione

Integra i test di resilienza introducendo consapevolmente errori nel sistema per misurarne la capacità in caso di scenari destabilizzanti. I test di resilienza, diversamente dai test funzionali e dagli unit test che di solito sono integrati nei cicli di implementazione, si concentrano sull'identificazione di errori imprevisti nel sistema. Puoi iniziare l'integrazione dei test di resilienza nella fase di pre-produzione, ma stabilisci l'obiettivo di implementare questi test in produzione durante le giornate di gioco.

Risultato desiderato: maggiore fiducia nella capacità del sistema di resistere al degrado nella produzione grazie ai test di resilienza. Gli esperimenti identificano i punti di debolezza che potrebbero causare errori, consentendoti di migliorare il sistema per mitigare automaticamente ed efficacemente errori e danneggiamento.

Anti-pattern comuni:

- Mancanza di osservabilità e monitoraggio nei processi di implementazione.
- Dipendenza dagli esseri umani per risolvere gli errori del sistema.
- Meccanismi di analisi di scarsa qualità.
- Supporto per i problemi noti del sistema e mancanza di sperimentazione per identificare eventuali incognite.
- Identificazione degli errori, ma nessuna risoluzione.
- Nessuna documentazione degli esiti e dei runbook.

Vantaggi dell'adozione delle best practice: i test di resilienza integrati nelle implementazioni consentono di identificare problemi sconosciuti nel sistema che altrimenti passerebbero inosservati, con conseguenti tempi di inattività nella produzione. L'identificazione di queste incognite nel sistema ti consente di documentare gli esiti, integrare i test nel processo CI/CD e creare runbook che semplificano la mitigazione attraverso meccanismi efficienti e ripetibili.

Livello di rischio associato se questa best practice non fosse adottata: medio

Guida all'implementazione

I moduli di test di resilienza più comuni che possono essere integrati nelle implementazioni del sistema sono il ripristino di emergenza e l'ingegneria del caos.

- Includi aggiornamenti ai piani di disaster recovery e alle procedure operative standard (SOPs) in qualsiasi implementazione significativa.
- Integra i test di affidabilità nelle pipeline di implementazione automatizzate. Servizi come AWS Resilience Hub possono essere integrati nella pipeline CI/CD al fine di valutare in modo continuo e automatico la resilienza nell'ambito di ogni implementazione.
- Definisci le tue applicazioni in AWS Resilience Hub. Le valutazioni della resilienza generano frammenti di codice che ti aiutano a creare procedure di ripristino come documenti di AWS Systems Manager per le tue applicazioni e forniscono un elenco di monitor e allarmi Amazon CloudWatch consigliati.

- Una volta aggiornati i piani di disaster recovery e SOPs i tuoi piani di disaster recovery, completa i test di disaster recovery per verificarne l'efficacia. I test di ripristino di emergenza consentono di determinare se è possibile ripristinare il sistema dopo un evento e tornare alle normali operazioni. Puoi simulare varie strategie di ripristino di emergenza e determinare se la pianificazione è sufficiente a soddisfare i requisiti di operatività. Le strategie di ripristino di emergenza più comuni includono backup e ripristino, pilot light, cold standby, warm standby, standby a caldo e attivo-attivo e si differenziano tutte per costi e complessità. Prima dei test di disaster recovery, ti consigliamo di definire il Recovery Time Objective (RTO) e il Recovery Point Objective (RPO) per semplificare la scelta della strategia da simulare. AWS offre strumenti di disaster recovery AWS Elastic Disaster Recoveryper aiutarvi a iniziare con la pianificazione e i test.
- Gli esperimenti di ingegneria del caos introducono interruzioni nel sistema, come interruzioni di rete ed errori del servizio. Simulando con gli errori controllati, puoi scoprire le vulnerabilità del sistema contenendo al contempo l'impatto degli errori inseriti. Analogamente alle altre strategie, esegui simulazioni controllate di guasti in ambienti non di produzione, con servizi come AWS Fault Injection Service, per acquisire sicurezza prima dell'implementazione in produzione.

Risorse

Documenti correlati:

- Experiment with failure using resilience testing to build recovery preparedness
- Valutazione continua della resilienza delle applicazioni con e AWS Resilience HubAWS CodePipeline
- Architettura di disaster recovery (DR) attiva AWS, parte 1: Strategie per il ripristino nel cloud
- Verify the resilience of your workloads using Chaos Engineering
- Principles of Chaos Engineering
- Workshop su Chaos Engineering

Video correlati:

- AWS re:Invent 2020: Testing Resilience using Chaos Engineering
- Migliora la resilienza delle applicazioni con il servizio AWS Fault Injection
- Prepara e proteggi le tue applicazioni dalle interruzioni con AWS Resilience Hub

REL08-BP04 Implementazione utilizzando un'infrastruttura immutabile

L'infrastruttura immutabile è un modello che richiede che non vengano applicati aggiornamenti, patch di sicurezza o modifiche di configurazione sui carichi di lavoro di produzione. Quando è necessaria una modifica, l'architettura viene costruita su una nuova infrastruttura e distribuita alla produzione.

Segui una strategia di implementazione dell'infrastruttura immutabile per aumentare l'affidabilità, la coerenza e la riproducibilità nelle implementazioni dei carichi di lavoro.

Risultato desiderato: con un'infrastruttura immutabile, non sono consentite modifiche locali (in-place) per l'esecuzione delle risorse dell'infrastruttura all'interno di un carico di lavoro. Invece, quando è necessaria una modifica, un nuovo set di risorse infrastrutturali aggiornate contenente tutte le modifiche necessarie viene implementato in parallelo alle risorse esistenti. Questa implementazione viene convalidata automaticamente e, in caso di successo, il traffico viene gradualmente trasferito al nuovo set di risorse.

Questa strategia di implementazione si applica, ad esempio, agli aggiornamenti software, alle patch di sicurezza, alle modifiche apportate all'infrastruttura, agli aggiornamenti della configurazione e agli aggiornamenti delle applicazioni.

Anti-pattern comuni:

- Implementazione locale (in-place) di modifiche alle risorse dell'infrastruttura in esecuzione.

Vantaggi dell'adozione di questa best practice:

- Maggiore coerenza tra gli ambienti: l'assenza di differenze nelle risorse dell'infrastruttura tra ambienti garantisce l'aumento della coerenza e la semplificazione dei test.
- Riduzione delle deviazioni di configurazione: sostituendo le risorse dell'infrastruttura con una configurazione nota e controllata dalla versione, l'infrastruttura viene impostata a uno stato noto, testato e affidabile, evitando deviazioni di configurazione.
- Implementazioni atomiche affidabili: il completamento delle implementazioni avviene con successo o non cambia nulla, così da aumentare coerenza e affidabilità del processo di implementazione.
- Implementazioni semplificate: le implementazioni sono semplificate poiché non devono supportare gli aggiornamenti. Gli aggiornamenti sono solo nuove implementazioni.
- Implementazioni più sicure con processi di rollback e ripristino rapidi: le implementazioni sono più sicure poiché la versione funzionante precedente non viene modificata. Puoi eseguire il rollback se vengono rilevati errori.

- Livello di sicurezza migliorato: non consentendo modifiche all'infrastruttura, i meccanismi di accesso remoto (ad esempio) possono essere disabilitati. SSH Questo riduce il vettore di attacco, migliorando il profilo di sicurezza dell'organizzazione.

Livello di rischio associato se questa best practice non fosse adottata: medio

Guida all'implementazione

Automazione

Nel definire una strategia di implementazione dell'infrastruttura immutabile, si consiglia di utilizzare il più possibile l'automazione per aumentare la riproducibilità e ridurre al minimo i potenziali errori umani. Per maggiori dettagli, consulta REL08-BP05 Implementare le modifiche con l'automazione e Automatizzare distribuzioni sicure e pratiche.

Con il modello Infrastructure as code (IaC), le fasi di provisioning, orchestrazione e implementazione dell'infrastruttura sono definite in modo programmatico, descrittivo e dichiarativo e archiviate in un sistema di controllo del codice sorgente. L'utilizzo del modello Infrastructure as code (IaC) semplifica l'automazione dell'implementazione dell'infrastruttura e aiuta a raggiungere l'immutabilità dell'infrastruttura.

Modelli di implementazione

Quando è richiesta una modifica del carico di lavoro, la strategia di implementazione immutabile dell'infrastruttura impone l'implementazione di un nuovo set di risorse dell'infrastruttura, comprese tutte le modifiche necessarie. È importante che questo nuovo set di risorse si basi su un modello di implementazione che riduca al minimo l'impatto sugli utenti. Esistono due strategie principali per questa implementazione:

Distribuzione canary: è la pratica di indirizzare un piccolo numero di clienti alla nuova versione, in genere in esecuzione su una singola istanza di servizio (la release canary). Quindi analizzerai in modo approfondito le modifiche di comportamento o gli errori generati. Puoi rimuovere il traffico dalla release canary in caso di problemi critici e reindirizzare gli utenti alla versione precedente. Se l'implementazione ha esito positivo, puoi continuare a implementare alla velocità desiderata, monitorando al contempo le modifiche per individuare eventuali errori, fino al completamento dell'implementazione. AWS CodeDeploy può essere configurato con una configurazione di distribuzione che consenta una distribuzione canaria.

Implementazione blu/verde: simile alla distribuzione canary, tranne per il fatto che un intero parco dell'applicazione è implementato in parallelo. Puoi alternare le implementazioni tra i due stack

(blu e verde). Ancora una volta, puoi inviare il traffico alla nuova versione e tornare alla versione precedente in caso di problemi con l'implementazione. Di solito tutto il traffico viene trasferito contemporaneamente, tuttavia puoi anche utilizzare frazioni del traffico verso ciascuna versione per accelerare l'adozione della nuova versione utilizzando le DNS funzionalità di routing ponderato di Amazon Route 53. AWS CodeDeploy e AWS Elastic Beanstalkpuò essere configurato con una configurazione di distribuzione che consente una distribuzione blu/verde.

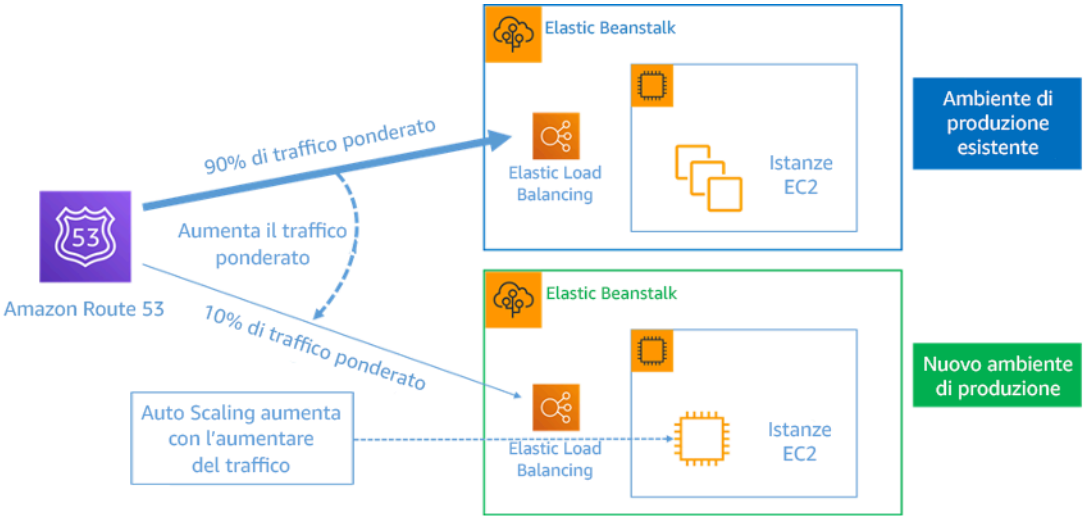


Figura 8: Implementazione blu/verde con AWS Elastic Beanstalk e Amazon Route 53

Rilevamento delle deviazioni

Per deviazione si intende qualsiasi modifica che causa uno stato o una configurazione di una risorsa dell'infrastruttura diversi da quelli previsti. Qualsiasi tipo di modifica non gestita della configurazione è contraria al concetto di infrastruttura immutabile e tale modifica dovrebbe essere individuata e corretta per implementare con successo l'infrastruttura immutabile.

Passaggi dell'implementazione

- Non autorizzare la modifica locale (in-place) delle risorse dell'infrastruttura in esecuzione.
  - È possibile utilizzare AWS Identity and Access Management (IAM) per specificare chi o cosa può accedere a servizi e risorse AWS, gestire centralmente le autorizzazioni dettagliate e analizzare l'accesso per perfezionare le autorizzazioni in tutto il mondo. AWS
- Automatizza l'implementazione delle risorse dell'infrastruttura per aumentare la riproducibilità e ridurre al minimo i potenziali errori umani.
  - Come descritto nel AWS white paper Introduzione a DevOps on, l'automazione è un elemento fondamentale per i servizi ed è supportata internamente in tutti AWS i servizi, le funzionalità e le offerte.

- La prepreparazione di Amazon Machine Image (AMI) può velocizzare i tempi di avvio. EC2Image Builder è un AWS servizio completamente gestito che consente di automatizzare la creazione, la manutenzione, la convalida, la condivisione e l'implementazione di applicazioni personalizzate, sicure e personalizzate per up-to-date Linux o Windows. AMI
- Alcuni dei servizi che supportano l'automazione sono:
  - AWS Elastic Beanstalkè un servizio per distribuire e scalare rapidamente applicazioni web sviluppate con Java,. NET, Node.jsPHP, Python, Ruby, Go e Docker su server familiari come Apache, NGINX Passenger e. IIS
  - AWS Protonaiuta i team della piattaforma a connettere e coordinare tutti i diversi strumenti necessari ai team di sviluppo per il provisioning dell'infrastruttura, la distribuzione del codice, il monitoraggio e gli aggiornamenti. AWS Proton abilita l'infrastruttura automatizzata come il provisioning del codice e la distribuzione di applicazioni serverless e basate su container.
- L'utilizzo dell'infrastruttura come codice semplifica l'automazione dell'implementazione dell'infrastruttura e aiuta a raggiungere l'immutabilità dell'infrastruttura. AWS fornisce servizi che consentono la creazione, l'implementazione e la manutenzione dell'infrastruttura in modo programmatico, descrittivo e dichiarativo.
  - AWS CloudFormationaiuta gli sviluppatori a creare AWS risorse in modo ordinato e prevedibile. Le risorse sono scritte in file di testo utilizzando il formato JSON oYAML. I modelli richiedono una sintassi e una struttura specifiche che dipendono dai tipi di risorse create e gestite. È possibile creare le risorse in JSON o YAML con qualsiasi editor di codice AWS Cloud9, ad esempio archiviarle in un sistema di controllo delle versioni e quindi CloudFormation creare i servizi specificati in modo sicuro e ripetibile.
  - AWS Serverless Application Model (AWS SAM) è un framework open source su cui è possibile creare applicazioni serverless. AWS AWS SAM si integra con altri AWS servizi ed è un'estensione di. AWS CloudFormation
  - AWS Cloud Development Kit (AWS CDK) è un framework di sviluppo software open source per modellare ed eseguire il provisioning delle risorse delle applicazioni cloud utilizzando linguaggi di programmazione familiari. È possibile utilizzare AWS CDK per modellare l'infrastruttura delle applicazioni utilizzando TypeScript Python, Java e. NET. AWS CDK utilizza AWS CloudFormation in background per fornire risorse in modo sicuro e ripetibile.
  - AWS Cloud Control APIintroduce un set comune di Create, Read, Update, Delete e List (CRUDL) APIs per aiutare gli sviluppatori a gestire la propria infrastruttura cloud in modo semplice e coerente. Cloud Control API Common APIs consente agli sviluppatori di gestire in modo uniforme il ciclo di vita dei servizi e di AWS terze parti.

- Applica modelli di implementazione che riducano al minimo l'impatto sugli utenti.
  - Distribuzione canary:
    - Configura una distribuzione di API Gateway Canary Release
    - Crea una pipeline con implementazioni Canary per Amazon utilizzando ECS AWS App Mesh
  - Implementazioni blu/verdi: il white paper Blue/Green Deployments on descrive esempi di tecniche per implementare strategie di implementazione blu/green. AWS
- Rileva le deviazioni a livello di configurazione o stato. Per ulteriori informazioni, consulta Detecting unmanaged configuration changes to stacks and resources.

Risorse

Best practice correlate:

- REL08-BP05 Implementa le modifiche con l'automazione

Documenti correlati:

- Automatizzazione di distribuzioni pratiche e sicure
- Sfruttare per creare un'infrastruttura immutabile AWS CloudFormation presso Nubank
- Infrastructure as code
- Implementazione di un allarme per rilevare automaticamente la deriva nelle pile AWS CloudFormation

Video correlati:

- AWS re:Invent 2020: affidabilità, coerenza e fiducia grazie all'immutabilità

REL08-BP05 Implementa le modifiche con l'automazione

Le implementazioni e l'applicazione di patch sono automatizzate per eliminare l'impatto negativo.

Apportare modifiche ai sistemi produttivi è una delle maggiori aree di rischio per molte organizzazioni. Riteniamo che le implementazioni siano un problema prioritario da risolvere insieme ai problemi aziendali affrontati dal software. Oggi, ciò significa l'uso dell'automazione ovunque sia pratica nelle operazioni, inclusi test e implementazione di modifiche, aggiunta o rimozione di capacità e migrazione dei dati.

Risultato desiderato: integrazione della sicurezza dell'implementazione automatizzata nel processo di rilascio con test di pre-produzione completi, rollback automatici e implementazioni di produzione scaglionate. Questa automazione riduce al minimo il potenziale impatto sulla produzione causato da implementazioni non riuscite e gli sviluppatori non devono più monitorare attivamente le implementazioni in produzione.

Anti-pattern comuni:

- Esegui le modifiche manualmente.
- Non esegui le fasi nell'automazione tramite flussi di lavoro manuali di emergenza.
- Non segui i piani e i processi stabiliti a favore di tempistiche accelerate.
- Esegui implementazioni successive rapide senza attendere il tempo di incorporamento.

Vantaggi dell'adozione di questa best practice: l'utilizzo dell'automazione per implementare tutte le modifiche elimina la possibilità di introdurre errori umani, oltre a offrire la possibilità di eseguire test prima di apportare modifiche alla produzione. L'esecuzione di questo processo prima del passaggio in produzione verifica che i piani siano completi. Inoltre, il rollback automatico del processo di rilascio può identificare i problemi di produzione e riportare il carico di lavoro allo stato operativo precedente.

Livello di rischio associato se questa best practice non fosse adottata: medio

Guida all'implementazione

Automatizzazione della pipeline di implementazione Le pipeline di implementazione permettono di richiamare test automatici, rilevare le anomalie e interrompere la pipeline a una determinata fase prima dell'implementazione in produzione o eseguire automaticamente il ripristino di una modifica. Parte integrante di ciò è l'adozione della cultura basata sull'integrazione continua e sulla consegna/implementazione continua (CI/CD), dove un commit o una modifica del codice passa lungo varie fasi automatizzate, dalle fasi di creazione e test, fino all'implementazione negli ambienti di produzione.

Anche se la prassi comune suggerisce di includere le persone nelle procedure operative più difficili, suggeriamo di automatizzare le procedure più difficili proprio per questo motivo.

Passaggi dell'implementazione

Per automatizzare le implementazioni ed eliminare le operazioni manuali, segui questi passaggi:

- Configura un repository di codice per archiviare il codice in modo sicuro: utilizza AWS CodeCommit, per la creazione di un repository sicuro basato su Git.

- Configura un servizio di integrazione continua per compilare il codice sorgente, eseguire test e creare artefatti di distribuzione: per configurare un progetto di compilazione a questo scopo, vedi Guida introduttiva all'utilizzo della console. AWS CodeBuild
- Configura un servizio di distribuzione che automatizzi le distribuzioni delle applicazioni e gestisca la complessità degli aggiornamenti delle applicazioni senza fare affidamento su distribuzioni manuali soggette a errori: AWS CodeDeployautomatizza le distribuzioni software su una varietà di servizi di elaborazione, come Amazon e i tuoi server locali. EC2 AWS FargateAWS Lambda Per configurare questi passaggi, consulta Guida introduttiva. CodeDeploy
- Imposta un servizio di distribuzione continua in grado di automatizzare le pipeline di rilascio per aggiornamenti più rapidi e affidabili delle applicazioni e dell'infrastruttura: prendi in considerazione l'utilizzo di AWS CodePipeline per automatizzare le tue pipeline di rilascio. Per maggiori dettagli, consulta i CodePipelinetutorial.

Risorse

Best practice correlate:

- OPS05-BP04 Usa sistemi di gestione della compilazione e dell'implementazione
- OPS05- BP1 0 Integrazione e implementazione completamente automatizzate
- OPS06-BP02 Implementazioni di test
- OPS06-BP04 Automatizza i test e il rollback

Documenti correlati:

- Distribuzione continua di pile annidate utilizzando AWS CloudFormationAWS CodePipeline
- CI/CD completo con AWS CodeCommit,, e AWS CodeBuildAWS CodeDeployAWS CodePipeline
- APNPartner: partner che possono aiutarti a creare soluzioni di implementazione automatizzate
- Marketplace AWS: prodotti per l'automazione delle implementazioni
- Automatizza i messaggi delle chat con webhook
- Amazon Builders' Library: garantire la sicurezza del rollback durante le distribuzioni
- Amazon Builders' Library: più velocità con una consegna continua
- Che cos'è AWS CodePipeline?
- Che cos'è CodeDeploy?

- AWS Systems Manager Patch Manager
- Che cos'è AmazonSES?
- What is Amazon Simple Notification Service?

Video correlati:

- AWS Summit 2019: CI/CD su AWS

Gestione dei guasti

Questions

- REL9. In che modo eseguire il backup dei dati?
- REL10. Come si utilizza l'isolamento dei guasti per proteggere il carico di lavoro?
- REL11. Come si progetta il carico di lavoro affinché resista ai guasti dei componenti?
- REL12. Come si testa l'affidabilità?
- REL13. Come si pianifica il ripristino di emergenza?

REL9. In che modo eseguire il backup dei dati?

Esegui il backup di dati, applicazioni e configurazione per soddisfare i requisiti relativi agli obiettivi dei tempi di ripristino (RTO) e agli obiettivi dei punti di ripristino (RPO).

Best practice

- REL09-BP01 Identifica ed esegui il backup di tutti i dati di cui è necessario eseguire il backup o riproduci i dati dalle fonti
- REL09-BP02 Backup sicuri e crittografati
- REL09-BP03 Esegui automaticamente il backup dei dati
- REL09-BP04 Eseguire il ripristino periodico dei dati per verificare l'integrità e i processi di backup

REL09-BP01 Identifica ed esegui il backup di tutti i dati di cui è necessario eseguire il backup o riproduci i dati dalle fonti

Scopri e utilizza le funzionalità di backup dei servizi e delle risorse di dati usati dal carico di lavoro. La maggior parte dei servizi offre funzionalità per eseguire il backup dei dati del carico di lavoro.

Risultato desiderato: le origini dati sono state identificate e classificate in base alla criticità. Quindi, stabilisci una strategia per il ripristino dei dati basata su. RPO Questa strategia prevede il backup di queste origini dati o la possibilità di riprodurre i dati da altre origini. In caso di perdita di dati, la strategia implementata consente il recupero o la riproduzione dei dati all'interno del territorio definito RPORTO.

Fase di maturità del cloud: di base

Anti-pattern comuni:

- Mancata conoscenza di tutte le origini dati per il carico di lavoro e della loro criticità.
- Non si eseguono backup delle origini dati critiche.
- Esecuzione di backup solo di alcune origini dati senza utilizzare la criticità come criterio.
- Nessuna frequenza definita RPO o di backup non può essere soddisfattaRPO.
- Nessuna valutazione della necessità di un backup o della possibilità di riprodurre i dati da altre origini.

Vantaggi dell'adozione di questa best practice: l'identificazione dei punti in cui sono necessari i backup e l'implementazione di un meccanismo per la creazione di backup, o la possibilità di riprodurre i dati da una fonte esterna, migliorano la capacità di ripristinare e recuperare i dati durante un'interruzione.

Livello di rischio associato se questa best practice non fosse adottata: elevato

Guida all'implementazione

Tutti gli archivi AWS dati offrono funzionalità di backup. Servizi come Amazon RDS e Amazon DynamoDB supportano inoltre il backup automatico che point-in-time consente il ripristino PITR (), che consente di ripristinare un backup in qualsiasi momento fino a cinque minuti o meno prima dell'ora corrente. Molti AWS servizi offrono la possibilità di copiare i backup su un altro. Regione AWS AWS Backup è uno strumento che offre la possibilità di centralizzare e automatizzare la protezione dei dati tra i servizi. AWS AWS Elastic Disaster Recoveryconsente di copiare carichi di lavoro completi sul server e mantenere una protezione continua dei dati da ambienti locali, inter-AZ o interregionali, con un Recovery Point Objective () misurato in secondi. RPO

Amazon S3 può essere utilizzato come destinazione di backup per sorgenti di dati autogestite e AWS gestite. AWS servizi come AmazonEBS, Amazon e Amazon RDS DynamoDB dispongono di funzionalità integrate per creare backup. È anche possibile utilizzare software di backup di terze parti.

È possibile eseguire il backup dei dati locali per l'utilizzo o. Cloud AWS AWS Storage GatewayAWS DataSync È possibile usare i bucket Amazon S3 per archiviare questi dati in AWS. Amazon S3 offre più livelli di archiviazione come Amazon S3 Glacier o S3 Glacier Deep Archive per ridurre i costi dell'archiviazione di dati.

Potresti essere in grado di soddisfare le esigenze di recupero dei dati riproducendo i dati da altre origini. Ad esempio, i nodi di ElastiCache replica Amazon o le repliche di RDS lettura di Amazon potrebbero essere utilizzati per riprodurre i dati in caso di perdita del primario. Nei casi in cui fonti come questa possono essere utilizzate per soddisfare i tuoi obiettivi Recovery Point Objective (RPO) e Recovery Time Objective (RTO), potresti non aver bisogno di un backup. Un altro esempio, se lavori con AmazonEMR, potrebbe non essere necessario eseguire il backup del tuo HDFS data store, purché sia possibile riprodurre i dati in Amazon EMR da Amazon S3.

Quando scegli una strategia di backup, devi considerare il tempo necessario per il ripristino dei dati. Il tempo necessario per il ripristino dei dati dipende dal tipo di backup (nel caso di una strategia di backup) o dalla complessità del meccanismo di riproduzione dei dati. Questo tempo dovrebbe rientrare nel carico RTO di lavoro.

Passaggi dell'implementazione

1. Identifica tutte le origini dati per il carico di lavoro. L'archiviazione dei dati può avvenire su varie risorse come database, volumi, file system, sistemi di log e storage a oggetti. Consulta la sezione Risorse per trovare i documenti correlati sui diversi AWS servizi in cui vengono archiviati i dati e sulle funzionalità di backup fornite da questi servizi.
2. Classifica le origini dati in base alla criticità. I diversi set di dati avranno diversi livelli di criticità per un carico di lavoro e quindi diversi requisiti di resilienza. Ad esempio, alcuni dati potrebbero essere critici e richiederne una quantità RPO prossima allo zero, mentre altri dati potrebbero essere meno critici e tollerare una perdita di dati maggiore RPO e parziale. Analogamente, anche set di dati diversi potrebbero avere RTO requisiti diversi.
3. Utilizza AWS i nostri servizi di terze parti per creare backup dei dati. AWS Backupè un servizio gestito che consente di creare backup di varie fonti di dati su. AWSAWS Elastic Disaster Recoverygestisce la replica automatica dei dati in meno di un secondo su un. Regione AWS La maggior parte AWS dei servizi dispone anche di funzionalità native per la creazione di backup. Marketplace AWS Ha molte soluzioni che forniscono anche queste funzionalità. Consulta la sezione Risorse più avanti per informazioni su come creare backup dei dati da vari servizi AWS .
4. Per i dati non sottoposti a backup, definisci un meccanismo di riproduzione dei dati. Puoi decidere di non eseguire il backup di dati riproducibili da altre origini per vari motivi. Potrebbe

essere più conveniente riprodurre i dati dalle origini, quando necessario, piuttosto che creare un backup, dato che l'archiviazione dei backup può comportare dei costi. Un altro esempio è il caso in cui il ripristino da un backup richiede più tempo rispetto alla riproduzione dei dati dalle fonti, con conseguente violazione. RTO In queste situazioni, è necessario considerare i compromessi e stabilire un processo ben definito per la riproduzione dei dati da queste origini quando è necessario il ripristino dei dati. Ad esempio, se hai caricato dati da Amazon S3 in un data warehouse (come Amazon Redshift) MapReduce o in un cluster (come EMR Amazon) per eseguire analisi su tali dati, questo potrebbe essere un esempio di dati che possono essere riprodotti da altre fonti. Finché i risultati di queste analisi sono archiviati da qualche parte o riproducibili, non si verificherebbe alcuna perdita di dati a causa di un guasto nel data warehouse o nel cluster. MapReduce Altri esempi che possono essere riprodotti dai sorgenti includono le cache (come Amazon ElastiCache) o le repliche di RDS lettura.

5. Definisci una cadenza per il backup dei dati. La creazione di backup delle fonti di dati è un processo periodico e la frequenza deve dipendere da. RPO

Livello di impegno per il piano di implementazione: moderato

Risorse

Best practice correlate:

REL13-BP01 Definire gli obiettivi di ripristino per tempi di inattività e perdita di dati

REL13-BP02 Utilizzare strategie di ripristino definite per raggiungere gli obiettivi di ripristino

Documenti correlati:

- Che cos'è? AWS Backup
- Che cos'è AWS DataSync?
- What is Volume Gateway?
- APNPartner: partner che possono aiutarti con il backup
- Marketplace AWS: prodotti che possono essere utilizzati per il backup
- EBSIstantanee Amazon
- Eseguire il backup di Amazon EFS
- Backup di Amazon FSx per Windows File Server
- Backup e ripristino ElastiCache per Redis

- Creating a DB Cluster Snapshot in Neptune
- Creating a DB Snapshot
- Creazione di una EventBridge regola che si attiva in base a una pianificazione
- Replica tra regioni con Amazon S3
- EFS-a- EFS AWS Backup
- Exporting Log Data to Amazon S3
- Gestione del ciclo di vita degli oggetti
- Backup e ripristino on demand per DynamoDB
- oint-in-timeRipristino P per DynamoDB
- Utilizzo delle istantanee OpenSearch di Amazon Service Index
- Che cos'è AWS Elastic Disaster Recovery?

Video correlati:

- AWS re:Invent 2021 - Backup, disaster recovery e protezione dal ransomware con AWS
- AWS Backup Demo: Backup su più account e più regioni
- AWS re:Invent 2019: Approfondimento su, ft. AWS Backup Spazio su rack () STG341

Esempi correlati:

- Well-Architected Lab - Implementazione della replica bidirezionale tra regioni () per Amazon S3 CRR
- Well-Architected Lab: esecuzione di test del backup e del ripristino di dati
- Well-Architected Lab: backup e ripristino con failback per un carico di lavoro di analisi
- Well-Architected Lab: ripristino di emergenza, backup e ripristino

REL09-BP02 Backup sicuri e crittografati

Controlla e rileva l'accesso ai backup utilizzando l'autenticazione e l'autorizzazione. Previeni e rileva se l'integrità dei dati dei backup è compromessa utilizzando la crittografia.

Anti-pattern comuni:

- Disporre di un accesso identico sia per i backup e l'automazione del ripristino sia per i dati.

- Non codificare i backup.

Vantaggi dell'adozione di questa best practice: la protezione dei backup previene la manomissione dei dati, mentre la crittografia dei dati impedisce l'accesso in caso di esposizione accidentale.

Livello di rischio associato se questa best practice non fosse adottata: elevato

Guida all'implementazione

Controlla e rileva l'accesso ai backup utilizzando l'autenticazione e l'autorizzazione, ad esempio (). AWS Identity and Access Management IAM Previeni e rileva se l'integrità dei dati dei backup è compromessa utilizzando la crittografia.

Amazon S3 supporta diversi metodi di crittografia dei dati a riposo. Utilizzando la crittografia lato server, Amazon S3 accetta anche dati non crittografati e li crittografa man mano che vengono memorizzati. Utilizzando la crittografia lato client, l'applicazione del carico di lavoro è responsabile della crittografia dei dati prima che vengano inviati ad Amazon S3. Entrambi i metodi consentono di utilizzare AWS Key Management Service (AWS KMS) per creare e archiviare la chiave dati, oppure di fornire la propria chiave, di cui l'utente è quindi responsabile. Utilizzando AWS KMS, puoi impostare politiche utilizzando IAM chi può e non può accedere alle tue chiavi di dati e ai dati decrittografati.

Per AmazonRDS, se hai scelto di crittografare i tuoi database, anche i tuoi backup vengono crittografati. I backup di DynamoDB sono sempre crittografati. Durante l'utilizzo AWS Elastic Disaster Recovery, tutti i dati in transito e a riposo vengono crittografati. Con Elastic Disaster Recovery, i dati inattivi possono essere crittografati utilizzando la chiave di EBS crittografia Amazon Encryption Volume Encryption predefinita o una chiave personalizzata gestita dal cliente.

Passaggi dell'implementazione

1. Utilizzo della crittografia su ciascuno dei datastore. Se i dati di origine sono crittografati, lo sarà anche il backup.
   - Usa la crittografia in AmazonRDS. . È possibile configurare la crittografia a riposo utilizzando AWS Key Management Service quando si crea un'RDSistanza.
   - Usa la crittografia sui EBS volumi Amazon. . Puoi configurare la crittografia predefinita o specificare una chiave univoca al momento della creazione del volume.
   - Utilizza la crittografia di Amazon DynamoDB necessaria. DynamoDB codifica tutti i dati a riposo. Puoi utilizzare una AWS KMS chiave AWS proprietaria o una KMS chiave AWS gestita, specificando una chiave archiviata nel tuo account.

- Crittografa i tuoi dati archiviati in Amazon EFS. Configura la crittografia al momento della creazione del file system.
- Configura la crittografia nelle regioni di origine e di destinazione. È possibile configurare la crittografia inattiva in Amazon S3 utilizzando chiavi archiviate inKMS, ma le chiavi sono specifiche della regione. Puoi specificare le chiavi di destinazione quando configuri la replica.
- Scegli se utilizzare la EBScrittografia Amazon predefinita o personalizzata per Elastic Disaster Recovery. Questa opzione esegue la crittografia dei dati a riposo replicati nei dischi della sottorete dell'area di staging e nei dischi replicati.

2. Implementazione delle autorizzazioni con privilegio minimo per accedere ai backup. Segui le best practice per limitare l'accesso a backup, snapshot e repliche in conformità con le best practice di sicurezza.

Risorse

Documenti correlati:

- Marketplace AWS: prodotti che possono essere utilizzati per il backup
- EBSCrittografia Amazon
- Amazon S3: protezione dei dati tramite la crittografia
- CRRConfigurazione aggiuntiva: replica di oggetti creati con crittografia lato server (SSE) utilizzando chiavi di crittografia archiviate in AWS KMS
- DynamoDB Encryption at Rest
- Crittografia delle risorse Amazon RDS
- Crittografia di dati e metadati in Amazon EFS
- Crittografia per i backup in AWS
- Gestione di tabelle crittografate
- Pilastro della sicurezza - AWS Well-Architected Framework
- Che cos'è? AWS Elastic Disaster Recovery

Esempi correlati:

- Well-Architected Lab - Implementazione della replica bidirezionale tra regioni () per Amazon S3 CRR

REL09-BP03 Esegui automaticamente il backup dei dati

Configura i backup in modo che vengano eseguiti automaticamente in base a una pianificazione periodica determinata dal Recovery Point Objective (RPO) o dalle modifiche nel set di dati. I set di dati critici con bassi requisiti di perdita di dati devono essere sottoposti a backup automatico su base frequente, mentre i dati meno critici, per i quali è accettabile una certa perdita, possono essere sottoposti a backup meno frequenti.

Risultato desiderato: un processo automatizzato che crea backup delle origini dati con una cadenza stabilita.

Anti-pattern comuni:

- Eseguire i backup manualmente.
- Utilizzare risorse che dispongono di funzionalità di backup, ma non includere il backup nell'automazione.

Vantaggi derivanti dall'adozione di questa best practice: l'automazione dei backup verifica che vengano eseguiti regolarmente in base alle tue RPO esigenze e avvisa l'utente se non vengono eseguiti.

Livello di rischio associato se questa best practice non fosse adottata: medio

Guida all'implementazione

AWS Backup può essere utilizzato per creare backup automatici di dati di varie fonti di dati. AWS Il backup RDS delle istanze Amazon può essere eseguito quasi ininterrottamente ogni cinque minuti e degli oggetti Amazon S3 può essere eseguito quasi ininterrottamente ogni quindici minuti, point-in-time prevedendo il ripristino PITR () in un momento specifico all'interno della cronologia di backup. Per altre fonti di AWS dati, come i EBS volumi Amazon, le tabelle Amazon DynamoDB o i file system FSx Amazon AWS Backup , è possibile eseguire backup automatici ogni ora. Questi servizi offrono anche funzionalità di backup native. AWS i servizi che offrono backup automatizzato con point-in-time ripristino includono Amazon DynamoDB, RDS Amazon e Amazon Keyspaces (per Apache Cassandra), che possono essere ripristinati in un momento specifico all'interno della cronologia di backup. La maggior parte degli altri servizi di archiviazione di dati AWS offre la possibilità di programmare backup periodici, anche ogni ora.

Amazon RDS e Amazon DynamoDB offrono backup continuo con ripristino. point-in-time Il controllo delle versioni di Amazon S3, una volta abilitato, è automatico. Amazon Data Lifecycle Manager può essere utilizzato per automatizzare la creazione, la copia e l'eliminazione di istantanee Amazon. EBS

Può anche automatizzare la creazione, la copia, la deprecazione e l'annullamento della registrazione di Amazon Machine Images () EBS AMIs supportate da Amazon e delle relative istantanee Amazon sottostanti. EBS

AWS Elastic Disaster Recovery fornisce una replica continua a livello di blocco dall'ambiente di origine (locale o) alla regione di ripristino di destinazione. AWS Point-in-timeLe EBS istantanee di Amazon vengono create e gestite automaticamente dal servizio.

Per una visualizzazione centralizzata dell'automazione e della cronologia dei backup, AWS Backup fornisce una soluzione di backup completamente gestita e basata su policy. Centralizza e automatizza il backup dei dati su più servizi AWS nel cloud e on-premises utilizzando AWS Storage Gateway.

Oltre al controllo delle versioni, Amazon S3 offre la funzionalità di replica. L'intero bucket S3 può essere replicato automaticamente in un altro bucket in una Regione AWS diversa.

Passaggi dell'implementazione

1. Identifica le origini dati al momento sottoposte a backup manuale. Per ulteriori dettagli, consulta REL09-BP01 Identifica ed esegui il backup di tutti i dati di cui è necessario eseguire il backup o riproduci i dati dalle fonti.
2. Determina il tipo di carico di RPO lavoro. Per ulteriori dettagli, consulta REL13-BP01 Definire gli obiettivi di ripristino per tempi di inattività e perdita di dati.
3. Utilizza una soluzione di backup automatizzata o un servizio gestito. AWS Backup è un servizio completamente gestito che semplifica la centralizzazione e l'automazione della protezione dei dati tra i AWS servizi, nel cloud e in locale. Usando piani di backup in AWS Backup, crea regole che definiscano le risorse di cui eseguire il backup e la frequenza di creazione dei backup. Questa frequenza dovrebbe essere informata secondo quanto stabilito nella RPO Fase 2. Per indicazioni pratiche su come creare backup automatici utilizzando AWS Backup, consulta Testing Backup and Restore of Data. Le funzionalità di backup native sono offerte dalla maggior parte dei AWS servizi che archiviano i dati. Ad esempio, RDS possono essere sfruttate per backup automatici con point-in-time recovery ()PITR.
4. Per le origini dati non supportate da una soluzione di backup automatico o da un servizio gestito, come le origini dati on-premises o le code di messaggi, è consigliabile utilizzare una soluzione di terze parti affidabile per creare backup automatici. In alternativa, è possibile creare l'automazione per eseguire questa operazione utilizzando o. AWS CLI SDKs Puoi utilizzare AWS Lambda Functions o AWS Step Functions definire la logica coinvolta nella creazione di un backup dei dati e utilizzare Amazon EventBridge per richiamarlo con una frequenza basata sulla tuaRPO.

Livello di impegno per il piano di implementazione: basso

Risorse

Documenti correlati:

- APNPartner: partner che possono aiutarti con il backup
- Marketplace AWS: prodotti che possono essere utilizzati per il backup
- Creazione di una EventBridge regola che si attiva in base a una pianificazione
- Che cos'è AWS Backup?
- Che cos'è AWS Step Functions?
- Che cos'è AWS Elastic Disaster Recovery?

Video correlati:

- AWS re:Invent 2019: Approfondimento su AWS Backup, ft. Spazio su rack () STG341

Esempi correlati:

- Well-Architected Lab: esecuzione di test del backup e del ripristino di dati

REL09-BP04 Eseguire il ripristino periodico dei dati per verificare l'integrità e i processi di backup

Verifica che l'implementazione del processo di backup soddisfi i Recovery Time Objectives (RTO) e Recovery Point Objectives (RPO) eseguendo un test di ripristino.

Risultato desiderato: i dati dei backup vengono ripristinati periodicamente utilizzando meccanismi ben definiti per verificare che il ripristino sia possibile entro l'obiettivo di tempo di ripristino stabilito (RTO) per il carico di lavoro. Verificate che il ripristino da un backup produca una risorsa contenente i dati originali senza che nessuno di essi sia danneggiato o inaccessibile e che la perdita dei dati rientri nell'obiettivo del punto di ripristino (). RPO

Anti-pattern comuni:

- Ripristino di un backup, ma senza eseguire query sui dati o recuperarli per verificare di poter usare il ripristino.
- Presupporre l'esistenza di un backup.

- Presupporre che il backup di un sistema sia pienamente operativo e che i dati possano essere recuperati da esso.
- Supponendo che il tempo necessario per il ripristino o il ripristino dei dati da un backup rientri nel limite del carico RTO di lavoro.
- Supponendo che i dati contenuti nel backup rientrino nel carico di lavoro RPO
- Ripristino in base alle esigenze, senza usare un runbook o seguire una procedura automatica prestabilita.

Vantaggi derivanti dall'adozione di questa procedura ottimale: il test del ripristino dei backup consente di verificare che i dati possano essere ripristinati quando necessario senza preoccuparsi che i dati possano mancare o essere danneggiati, che il ripristino e il ripristino siano possibili all'interno del RTO carico di lavoro e che qualsiasi perdita di dati rientri nel carico di lavoro. RPO

Livello di rischio associato se questa best practice non fosse adottata: medio

Guida all'implementazione

La verifica delle capacità di backup e ripristino aumenta la fiducia nella capacità di eseguire queste azioni durante un'interruzione. Ripristina periodicamente i backup in una nuova posizione ed esegui test per verificare l'integrità dei dati. Alcuni test comuni da eseguire consistono nel verificare se tutti i dati sono disponibili, non sono danneggiati, sono accessibili e se qualsiasi perdita di dati rientra nell'ambito del RPO carico di lavoro. Tali test possono anche aiutare a verificare se i meccanismi di ripristino sono sufficientemente veloci da soddisfare il carico di lavoro. RTO

In questo modo è possibile creare un ambiente di test e ripristinare i backup per valutarne le RPO funzionalità RTO ed eseguire test sul contenuto e sull'integrità dei dati. AWS

Inoltre, Amazon RDS e Amazon DynamoDB point-in-time consentono il ripristino (). PITR Utilizzando il backup continuo, puoi ripristinare il set di dati allo stato in cui si trovava in una data e un'ora specificate.

Se tutti i dati sono disponibili, non sono danneggiati, sono accessibili e qualsiasi perdita di dati rientra nel carico di RPO lavoro. Tali test possono anche aiutare a verificare se i meccanismi di ripristino sono sufficientemente veloci da soddisfare il carico di lavoro. RTO

AWS Elastic Disaster Recovery offre istantanee di point-in-time ripristino continuo dei volumi AmazonEBS. Man mano che i server di origine vengono replicati, point-in-time gli stati vengono cronizzati nel tempo in base alla politica configurata. Elastic Disaster Recovery verifica l'integrità di questi snapshot avviando istanze per scopi di test ed esercitazione senza reindirizzare il traffico.

Passaggi dell'implementazione

1. Identifica le origini dati di cui stai eseguendo il backup e dove sono archiviati i backup. Per le linee guida di implementazione, consulta REL09-BP01 Identifica ed esegui il backup di tutti i dati di cui è necessario eseguire il backup o riproduci i dati dalle fonti.
2. Definisci criteri per la convalida dei dati per ciascuna origine dati. Tipi di dati differenti avranno proprietà diverse che potrebbero richiedere meccanismi di convalida diversi. Considera il modo in cui potrebbero essere convalidati questi dati prima di poterli utilizzare in produzione. Alcuni modi comuni per convalidare i dati sono l'uso delle loro proprietà dei dati e del backup, come il tipo di dati, il formato, la somma di controllo, la dimensione o la combinazione di questi elementi con una logica di convalida personalizzata. Ad esempio, può trattarsi di un confronto dei valori di checksum tra la risorsa ripristinata e l'origine dati al momento della creazione del backup.
3. Stabilisci RTO e ripristina RPO i dati in base alla loro criticità. Per le linee guida di implementazione, consulta REL13-BP01 Definire gli obiettivi di ripristino per tempi di inattività e perdita di dati.
4. Valuta la capacità di ripristino. Esamina la tua strategia di backup e ripristino per capire se è in grado di soddisfare le tue esigenze RTO eRPO, se necessario, adattala. AWS Resilience Hub ti consente di valutare il tuo carico di lavoro. La valutazione valuta la configurazione dell'applicazione rispetto alla politica di resilienza e segnala se RPO gli obiettivi prefissati possono essere raggiuntiRTO.
5. Esegui un ripristino di test utilizzando i processi attualmente in uso in produzione per il ripristino dei dati. Questi processi dipendono dal modo in cui è stato eseguito il backup dell'origine dati iniziale, dal formato e dalla posizione di archiviazione del backup stesso o dalla riproduzione dei dati da altre fonti. Ad esempio, in caso di utilizzo di un servizio gestito, come AWS Backup, potrebbe essere semplice ripristinare il backup in una nuova risorsa. In caso di utilizzo di AWS Elastic Disaster Recovery , è possibile avviare un'esercitazione di ripristino.
6. Convalida il ripristino dei dati dalla risorsa ripristinata in base ai criteri stabiliti in precedenza per la convalida dei dati. I dati ripristinati e recuperati contengono il record o la voce più recente al momento del backup? Questi dati rientrano nel carico di lavoro RPO per il carico di lavoro?
7. Misura il tempo necessario per il ripristino e il ripristino e confrontalo con quello stabilitoRTO. Questo processo rientra nel carico RTO di lavoro? Ad esempio, confronta i timestamp dell'inizio del processo di ripristino e del completamento della convalida del ripristino per calcolare la durata del processo. Tutte le AWS API chiamate hanno una marcatura temporale e queste informazioni sono disponibili in. AWS CloudTrail Sebbene queste informazioni possano fornire dettagli sull'inizio del processo di ripristino, la logica di convalida dovrebbe registrare il timestamp finale del completamento della convalida. Se utilizzi un processo automatizzato, puoi sfruttare servizi

come Amazon DynamoDB per archiviare queste informazioni. Inoltre, molti AWS servizi forniscono una cronologia degli eventi che fornisce informazioni con data e ora quando si sono verificate determinate azioni. All'interno AWS Backup, le azioni di backup e ripristino sono denominate processi e tali processi contengono informazioni sul timestamp come parte dei relativi metadati, che possono essere utilizzati per misurare il tempo necessario per il ripristino e il ripristino.

8. Informa le parti interessate se la convalida dei dati fallisce o se il tempo necessario per il ripristino e il ripristino supera il tempo stabilito per il carico di lavoro. RTO Quando si implementa l'automazione per eseguire questa operazione, ad esempio in questo laboratorio, è possibile utilizzare servizi come Amazon Simple Notification Service (AmazonSNS) per inviare notifiche push come e-mail o SMS alle parti interessate. Questi messaggi possono anche essere pubblicati su applicazioni di messaggistica come Amazon Chime, Slack o Microsoft Teams o utilizzati per creare attività utilizzando Systems OpsItems Manager AWS. OpsCenter
9. Automatizza questo processo per eseguirlo periodicamente. Ad esempio, è AWS Step Functions possibile utilizzare servizi come AWS Lambda o una State Machine in per automatizzare i processi di ripristino e ripristino e Amazon EventBridge può essere utilizzato per richiamare periodicamente questo flusso di lavoro di automazione, come mostrato nel diagramma di architettura seguente. Scopri come automatizzare la convalida del ripristino dei dati con. AWS Backup Inoltre, questo Well-Architected lab fornisce un'esperienza pratica su come realizzare l'automazione di alcuni dei passaggi qui descritti.

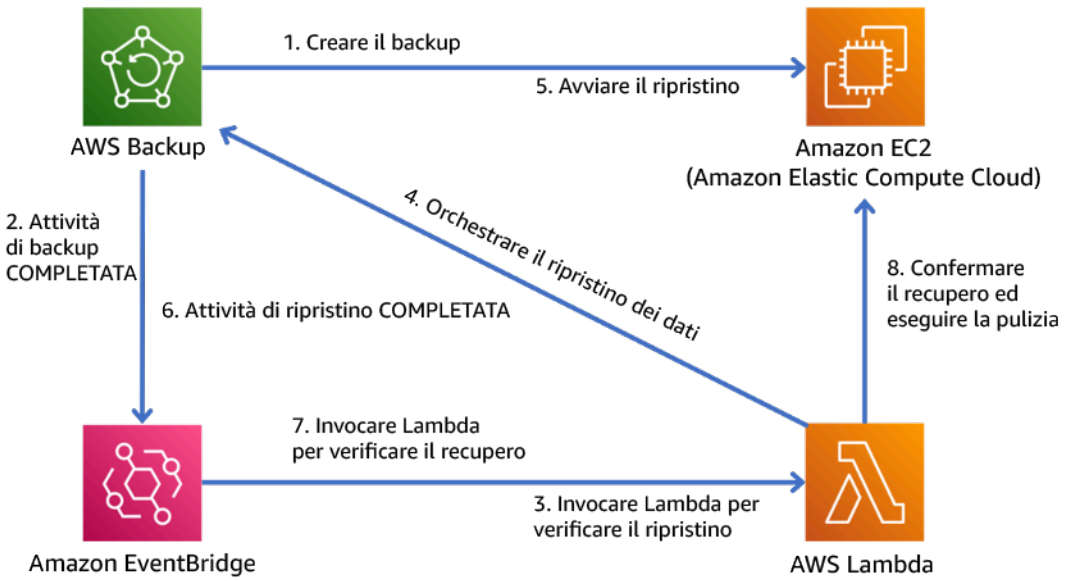


Figura 9. Processo di backup e ripristino automatico

Livello di impegno per il piano di implementazione: da moderato a elevato, in base alla complessità dei criteri di convalida.

Risorse

Documenti correlati:

- Automatizza la convalida del ripristino dei dati con AWS Backup
- APNPartner: partner che possono aiutarti con il backup
- Marketplace AWS: prodotti che possono essere utilizzati per il backup
- Creazione di una EventBridge regola che si attiva in base a una pianificazione
- Backup e ripristino on demand per DynamoDB
- Che cos'è AWS Backup?
- Che cos'è AWS Step Functions?
- Che cosa è AWS Elastic Disaster Recovery
- AWS Elastic Disaster Recovery

Esempi correlati:

- Well-Architected Lab: esecuzione di test del backup e del ripristino di dati

REL10. Come si utilizza l'isolamento dei guasti per proteggere il carico di lavoro?

Le barriere per l'isolamento dei guasti limitano l'effetto di un errore all'interno di un carico di lavoro a un numero limitato di componenti. I componenti al di fuori della barriera non subiscono gli effetti del guasto. Utilizzando più barriere per l'isolamento dei guasti, puoi limitare l'impatto sul carico di lavoro.

Best practice

- REL10-BP01 Implementazione del carico di lavoro in più sedi
- REL10-BP02 Seleziona le posizioni appropriate per l'implementazione in più sedi
- REL10-BP03 Ripristino automatico dei componenti vincolati a un'unica posizione
- REL10-BP04 Usa architetture a paratia per limitare l'ambito dell'impatto

REL10-BP01 Implementazione del carico di lavoro in più sedi

Distribuisci i dati e le risorse del carico di lavoro su più zone di disponibilità o, se necessario, su diverse Regioni AWS. Questi luoghi possono essere diversi a seconda delle necessità.

Uno dei principi fondamentali per la progettazione dei servizi AWS è la prevenzione di singoli punti di errore nell'infrastruttura fisica sottostante. Questo ci spinge a creare software e sistemi che utilizzano più zone di disponibilità e sono resistenti ai guasti di una singola zona. Allo stesso modo, i sistemi sono costruiti per resistere ai guasti di un singolo nodo di calcolo, singolo volume di archiviazione o singola istanza di un database. Quando si costruisce un sistema che si basa su componenti ridondanti, è importante assicurarsi che i componenti funzionino in modo indipendente e, nel caso di, in modo autonomo. Regioni AWS I vantaggi ottenuti dai calcoli di disponibilità teorica con componenti ridondanti sono validi solo se questo continua a essere vero.

Zone di disponibilità () AZs

Regioni AWS sono composte da più zone di disponibilità progettate per essere indipendenti l'una dall'altra. Ogni zona di disponibilità è separata da una distanza fisica significativa da altre zone per evitare scenari di guasto correlati, dovuti a rischi ambientali come incendi, inondazioni e tornado. Ogni zona di disponibilità ha anche un'infrastruttura fisica indipendente: connessioni dedicate di alimentazione di rete, fonti di alimentazione di backup autonome, servizi meccanici indipendenti e connettività di rete indipendente all'interno e all'esterno della zona di disponibilità. Questa struttura limita gli errori di uno qualsiasi di questi sistemi alla sola AZ interessata. Nonostante siano geograficamente separate, le zone di disponibilità sono situate nella stessa area regionale, il che consente una rete a elevato throughput e bassa latenza. L'intera area Regione AWS (in tutte le zone di disponibilità, costituita da più data center fisicamente indipendenti) può essere considerata come un unico obiettivo di implementazione logico per il carico di lavoro, inclusa la possibilità di replicare i dati in modo sincrono (ad esempio, tra database). Ciò ti consente di utilizzare le zone di disponibilità in una configurazione attiva/attiva o attiva/standby.

Le zone di disponibilità sono indipendenti e pertanto la disponibilità del carico di lavoro aumenta quando il carico di lavoro è progettato per utilizzare più zone di disponibilità. Alcuni AWS servizi (incluso il piano dati delle EC2 istanze Amazon) vengono distribuiti come servizi strettamente zonali in cui hanno condiviso il destino con la zona di disponibilità in cui si trovano. EC2Le istanze Amazon nell'altra versione non AZs saranno tuttavia interessate e continueranno a funzionare. Allo stesso modo, se un errore in una zona di disponibilità causa l'errore di un database Amazon Aurora, un'istanza Aurora di lettura-replica in una zona di disponibilità non interessata può essere automaticamente promossa a primaria. I servizi regionali AWS , come Amazon DynamoDB, utilizzano internamente più zone di disponibilità in una configurazione attiva/attiva per raggiungere gli

obiettivi di progettazione della disponibilità per quel servizio, senza che sia necessario configurare il posizionamento delle zone di disponibilità.

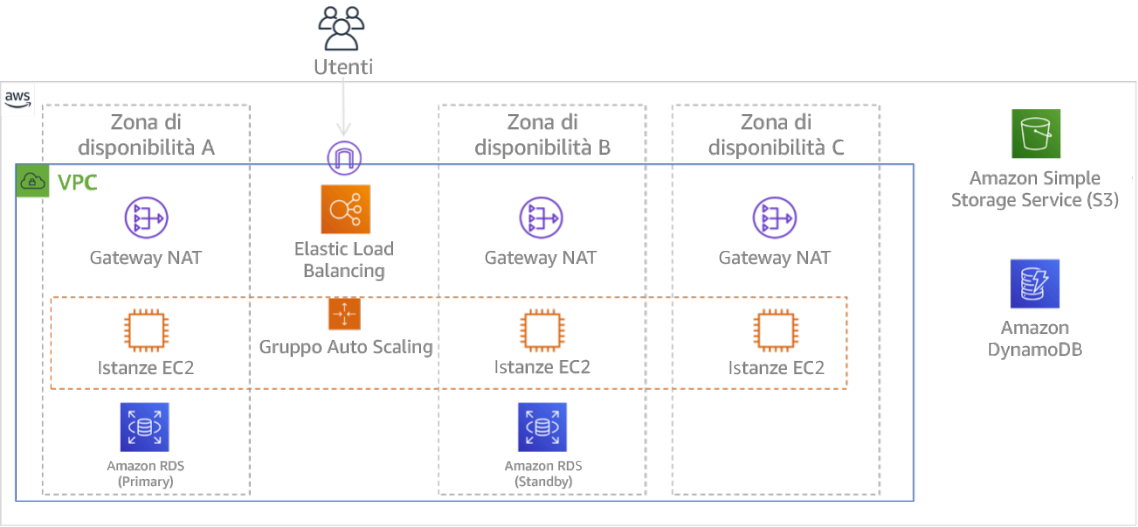


Figura 9: architettura multi-livello distribuita su tre zone di disponibilità. Nota: Amazon S3 e Amazon DynamoDB sono sempre ad AZ multiple automaticamente. ELBInoltre è distribuito in tutte e tre le zone.

Sebbene i piani di AWS controllo offrano in genere la possibilità di gestire le risorse all'interno dell'intera regione (più zone di disponibilità), alcuni piani di controllo (inclusi Amazon EC2 e AmazonEBS) hanno la capacità di filtrare i risultati in base a un'unica zona di disponibilità. Con questo approccio, la richiesta viene elaborata solo nella zona di disponibilità specificata, riducendo l'esposizione all'interruzione in altre zone di disponibilità. Questo AWS CLI esempio illustra come ottenere informazioni sulle EC2 istanze Amazon solo dalla zona di disponibilità us-east-2c:

```
AWS ec2 describe-instances --filters Name=availability-zone,Values=us-east-2c
```

AWS Zone locali

AWS Le Local Zone agiscono in modo simile alle Zone di disponibilità all'interno delle rispettive aree, Regione AWS in quanto possono essere selezionate come ubicazioni di collocamento per AWS risorse zonali come sottoreti e istanze. EC2 Ciò che le rende speciali è il fatto che non si trovano nelle aree associate Regione AWS, ma in prossimità di centri IT, industriali e popolati, dove oggi non esistono. Regione AWS Tuttavia, mantengono una connessione sicura e a larghezza di banda elevata tra i carichi di lavoro locali nella zona locale e quelli in esecuzione nella Regione AWS. È consigliabile utilizzare le zone locali AWS per implementare i carichi di lavoro più vicini agli utenti per requisiti a bassa latenza.

Amazon Global Edge Network

Amazon Global Edge Network è composto da posizioni edge in città di tutto il mondo. Amazon CloudFront utilizza questa rete per fornire contenuti agli utenti finali con una latenza inferiore. AWS Global Accelerator ti consente di creare endpoint di carico di lavoro in queste edge location per consentire l'onboarding sulla rete AWS globale vicino ai tuoi utenti. Amazon API Gateway consente API endpoint ottimizzati per l'edge utilizzando una CloudFront distribuzione per facilitare l'accesso dei clienti attraverso l'edge location più vicina.

Regioni AWS

Regioni AWS sono progettati per essere autonomi, pertanto, per utilizzare un approccio multiregionale, è necessario distribuire copie dedicate dei servizi in ciascuna regione.

Un approccio multiregionale è comune per le strategie di ripristino di emergenza volte al raggiungimento degli obiettivi di ripristino in caso di eventi isolati su larga scala. Consulta Pianificazione per il disaster recovery (DR) per ulteriori informazioni su queste strategie. Qui, tuttavia, ci concentriamo invece sulla disponibilità, che cerca di fornire un obiettivo medio di operatività nel tempo. Per gli obiettivi di alta disponibilità, un'architettura multi-regione sarà generalmente progettata per essere attiva/attiva, dove ogni copia del servizio (nelle rispettive regioni) è attiva (serve le richieste).

> **Raccomandazione**
>
> Gli obiettivi di disponibilità per la maggior parte dei carichi di lavoro possono essere soddisfatti utilizzando una strategia multi-AZ all'interno di una singola Regione AWS. Considera le architetture multi-regione solo quando i carichi di lavoro hanno requisiti di disponibilità estremi o altri obiettivi aziendali che richiedono un'architettura multi-regione.

AWS ti offre le funzionalità per gestire servizi in più regioni. Ad esempio, AWS fornisce la replica continua e asincrona dei dati utilizzando Amazon Simple Storage Service (Amazon S3) Replication, Amazon Read Replicas (inclusa Aurora Read Replicas) e RDS Amazon DynamoDB Global Tables. Con la replica continua, le versioni dei dati sono disponibili per un uso quasi immediato in ogni regione attiva.

In questo modo AWS CloudFormation, puoi definire la tua infrastruttura e distribuirla in modo uniforme su e indietro. Account AWS Regioni AWS Inoltre, AWS CloudFormation StackSets estende questa funzionalità consentendoti di creare, aggiornare o eliminare AWS CloudFormation pile su

più account e regioni con un'unica operazione. Per le distribuzioni di EC2 istanze Amazon, viene utilizzata una (AMIAmazon Machine Image) per fornire informazioni come la configurazione hardware e il software installato. Puoi implementare una pipeline Amazon EC2 Image Builder che crea ciò di cui AMIs hai bisogno e copiarlo nelle tue regioni attive. Ciò garantisce che questi Golden AMIs abbiano tutto ciò di cui hai bisogno per distribuire e scalare il tuo carico di lavoro in ogni nuova regione.

Per indirizzare il traffico, sia Amazon Route 53 che AWS Global Accelerator consentono la definizione di policy che determinano quali utenti accedono a quale endpoint regionale attivo. Con Global Accelerator imposti un valore di traffico per controllare la percentuale di traffico diretta a ciascun endpoint dell'applicazione. Route 53 supporta questo approccio percentuale e anche molte altre policy disponibili, tra cui quelle basate sulla geoprossimità e sulla latenza. Global Accelerator sfrutta automaticamente l'ampia rete di server AWS perimetrali per integrare il traffico verso la dorsale della AWS rete il prima possibile, con conseguente riduzione delle latenze di richiesta.

Tutte queste capacità operano in modo da preservare l'autonomia di ogni regione. Esistono pochissime eccezioni a questo approccio, inclusi i nostri servizi che forniscono una distribuzione edge globale (come Amazon CloudFront e Amazon Route 53), insieme al piano di controllo per il servizio AWS Identity and Access Management (IAM). La maggior parte dei servizi opera interamente all'interno di una singola regione.

Data center on-premises

Per i carichi di lavoro eseguiti in un data center locale, progetta un'esperienza ibrida quando possibile. AWS Direct Connect fornisce una connessione di rete dedicata dalla tua sede per AWS consentirti di funzionare in entrambi.

Un'altra opzione è quella di eseguire AWS l'infrastruttura e i servizi in locale utilizzando AWS Outposts. AWS Outposts è un servizio completamente gestito che estende AWS l'infrastrutturaAPIs, AWS i servizi e gli strumenti al data center. La stessa infrastruttura hardware utilizzata in Cloud AWS è installata nel data center. AWS Outposts vengono quindi collegati al più vicino Regione AWS. È quindi possibile utilizzarli AWS Outposts per supportare carichi di lavoro con bassa latenza o requisiti di elaborazione locale dei dati.

Livello di rischio associato se questa best practice non fosse adottata: elevato

Guida all'implementazione

- Utilizza più zone di disponibilità e. Regioni AWS Distribuisci i dati e le risorse del carico di lavoro su più zone di disponibilità o, se necessario, su diverse Regioni AWS. Questi luoghi possono essere diversi a seconda delle necessità.

- I servizi regionali sono implementati intrinsecamente in zone di disponibilità.
  - Ciò include Amazon S3, Amazon DynamoDB AWS Lambda e (quando non è connesso a) VPC
- Implementa il container, l'istanza e i carichi di lavoro basati su funzioni in più zone di disponibilità. Utilizza datastore multi-zona, inclusi sistemi di cache. Usa le funzionalità di Amazon EC2 Auto Scaling, il posizionamento delle ECS attività di Amazon, AWS Lambda la configurazione delle funzioni durante l'esecuzione nei tuoi e VPC ElastiCache i cluster.
  - Utilizza sottoreti che sono in zone di disponibilità separate nella distribuzione di gruppi Auto Scaling.
    - Esempio: distribuzione di istanze tra le zone di disponibilità
    - Scelta di regioni e zone di disponibilità
  - Utilizza i parametri di posizionamento delle ECS attività, specificando i gruppi di sottoreti del DB.
    - Strategie di collocamento delle ECS attività di Amazon
  - Usa sottoreti in più zone di disponibilità quando configuri una funzione da eseguire in. VPC
    - Configurazione di una AWS Lambda funzione per accedere alle risorse in Amazon VPC
  - Utilizza più zone di disponibilità con ElastiCache cluster.
    - Scelta di regioni e zone di disponibilità
- Se il carico di lavoro deve essere implementato in più regioni, scegli una strategia multi-regione. La maggior parte delle esigenze di affidabilità può essere soddisfatta in un'unica soluzione Regione AWS utilizzando una strategia a più zone di disponibilità. Quando necessario, utilizza una strategia multi-regione per soddisfare le tue esigenze aziendali.
  - AWS re:Invent 2018: modelli di architettura per applicazioni Active-Active in più regioni (09-R2) ARC2
    - Il backup su un altro Regione AWS può aggiungere un ulteriore livello di garanzia che i dati siano disponibili quando necessario.
    - Alcuni carichi di lavoro hanno requisiti normativi che prevedono l'utilizzo di una strategia multi-regione.
- Valuta AWS Outposts il tuo carico di lavoro. Se il carico di lavoro richiede bassa latenza nel data center on-premises o applica i requisiti di elaborazione dei dati locali, Quindi esegui AWS l'infrastruttura e i servizi in locale utilizzando AWS Outposts
  - Che cos'è AWS Outposts?

- Determina se AWS Local Zones ti aiuta a fornire servizi ai tuoi utenti. Se hai requisiti di bassa latenza, verifica se AWS Local Zones si trova vicino ai tuoi utenti. Se sì, utilizzale per implementare carichi di lavoro più vicini a tali utenti.
  - Zone locali AWS FAQ

Risorse

Documenti correlati:

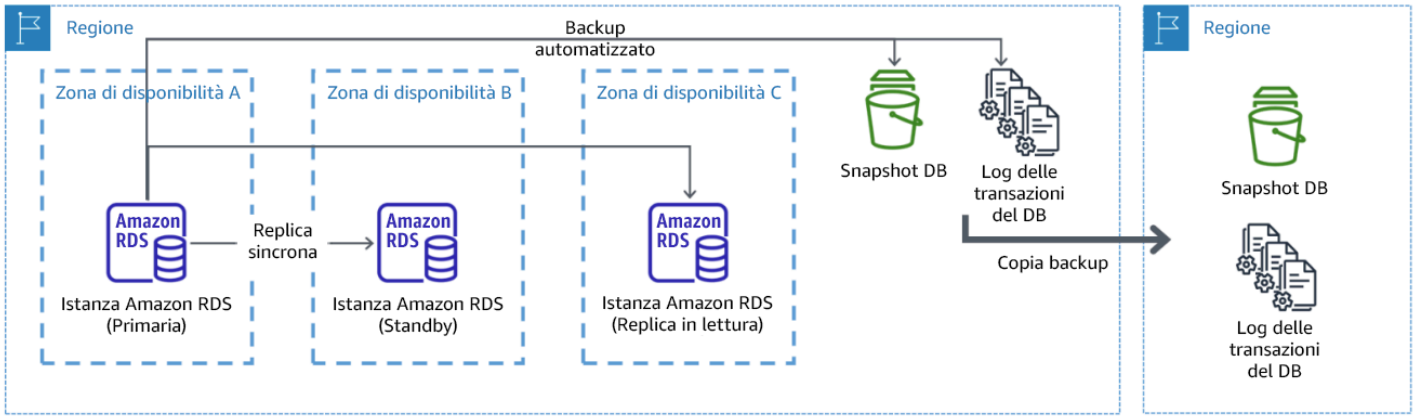
- Infrastruttura globale di AWS
- Zone locali AWS FAQ
- Strategie di collocamento delle ECS attività di Amazon
- Scelta di regioni e zone di disponibilità
- Esempio: distribuzione di istanze tra le zone di disponibilità
- Global Tables: Multi-Region Replication with DynamoDB
- Using Amazon Aurora global databases
- Serie di blog sulla creazione di un'applicazione multiregionale con AWS servizi
- Che cos'è AWS Outposts?

Video correlati:

- AWS re:Invent 2018: modelli di architettura per applicazioni Active-Active in più regioni (09-R2) ARC2
- AWS re:Invent 2019: Innovazione e funzionamento dell'infrastruttura di rete globale () AWS NET339

REL10-BP02 Seleziona le posizioni appropriate per l'implementazione in più sedi

Risultato desiderato: per un'elevata disponibilità, distribuisci sempre (quando possibile) i componenti del carico di lavoro in più zone di disponibilità (). AZs Per i carichi di lavoro con requisiti di resilienza estremi, valuta attentamente le opzioni per un'architettura multiregione.



implementazione resiliente di un database multi-AZ con backup in un'altra regione AWS

Anti-pattern comuni:

- Scelta di progettare un'architettura multi-regione quando un'architettura multi-AZ soddisferebbe i requisiti.
- Non si tiene conto delle dipendenze tra i componenti dell'applicazione se i requisiti di resilienza e multi-sede differiscono tra questi componenti.

Vantaggi dell'adozione di questa best practice: per ottenere resilienza, ricorri a un approccio che crei livelli di difesa. Un livello protegge da interruzioni più piccole e più comuni creando un'architettura ad alta disponibilità che utilizza più architetture. AZs Un altro livello di difesa è destinato a proteggere da eventi rari come disastri naturali diffusi e interruzioni a livello regionale. Questo secondo livello prevede l'architettura dell'applicazione in modo che si estenda su più fronti. Regioni AWS

- La differenza tra una disponibilità del 99,5% e una del 99,99% è di oltre 3,5 ore al mese. La disponibilità prevista di un carico di lavoro può raggiungere solo «quattro nove» se è suddivisa in più di un carico di lavoro. AZs
- Eseguendo il carico di lavoro su più livelliAZs, è possibile isolare i guasti di alimentazione, raffreddamento e rete e la maggior parte dei disastri naturali come incendi e inondazioni.
- L'implementazione di una strategia multi-regione per il tuo carico di lavoro aiuta a proteggerlo da disastri naturali diffusi che colpiscono un'ampia regione geografica di un paese o da guasti tecnici di portata regionale. Tieni presente che l'implementazione di un'architettura multi-regione può essere molto complessa e di solito non è necessaria per la maggior parte dei carichi di lavoro.

Livello di rischio associato se questa best practice non fosse adottata: elevato

Guida all'implementazione

Per un evento di emergenza basato sull'interruzione o sulla perdita parziale di una zona di disponibilità, l'implementazione di un carico di lavoro ad alta disponibilità in più zone di disponibilità all'interno di un'unica zona Regione AWS aiuta a mitigare i disastri naturali e tecnici. Ogni Regione AWS è composta da più zone di disponibilità, ciascuna isolata dagli errori nelle altre zone e separate da una distanza significativa. Tuttavia, per un evento di emergenza che include il rischio di perdere più componenti della zona di disponibilità, che si trovano a una distanza significativa l'uno dall'altro, è necessario implementare opzioni di ripristino di emergenza per mitigare gli errori di portata regionale. Per i carichi di lavoro che richiedono un'estrema resilienza (infrastrutture critiche, applicazioni sanitarie, infrastrutture di sistemi finanziari e così via), può essere necessaria una strategia multi-regione.

Passaggi dell'implementazione

1. Valuta il carico di lavoro e determina se le esigenze di resilienza possono essere soddisfatte con un approccio Multi-AZ (singolo Regione AWS) o se richiedono un approccio multiregionale. L'implementazione di un'architettura multi-regione per soddisfare questi requisiti introdurrà un'ulteriore complessità, quindi considera attentamente il tuo caso d'uso e i suoi requisiti. I requisiti di resilienza possono quasi sempre essere soddisfatti utilizzando un singolo Regione AWS. Per stabilire se è necessario utilizzare più regioni, considera i seguenti possibili requisiti:
   a. Disaster recovery (DR): per un evento di emergenza basato sull'interruzione o la perdita parziale di una zona di disponibilità, l'implementazione di un carico di lavoro ad alta disponibilità in più zone di disponibilità all'interno di una singola Regione AWS aiuta a mitigare i disastri naturali e tecnici. In caso di eventi di emergenza che comportano il rischio di perdere più componenti delle zone di disponibilità, che si trovano a una distanza significativa l'uno dall'altro, è necessario implementare il ripristino di emergenza in più regioni per mitigare i disastri naturali o gli errori tecnici di portata regionale.
   b. Alta disponibilità (HA): è possibile utilizzare un'architettura multiregionale (che ne utilizza più AZs in ogni regione) per ottenere una disponibilità superiore a quattro 9 (**>** 99,99%).
   c. Localizzazione degli stack: quando si distribuisce un carico di lavoro a un pubblico globale, è possibile distribuire stack localizzati in diversi gruppi di destinatari in quelle regioni. Regioni AWS La localizzazione può includere la lingua, la valuta e i tipi di dati memorizzati.
   d. Vicinanza agli utenti: quando si distribuisce un carico di lavoro a un pubblico globale, è possibile ridurre la latenza distribuendo gli stack vicino a dove si trovano gli utenti finali. Regioni AWS
   e. Residenza dei dati: alcuni carichi di lavoro sono soggetti a requisiti di residenza dei dati, in base ai quali i dati di determinati utenti devono rimanere all'interno dei confini di un determinato

Paese. In base alla normativa in questione, puoi scegliere di distribuire un intero stack, o solo i dati, all'interno di tali confini. Regione AWS

2. Ecco alcuni esempi di funzionalità multi-AZ fornite dai servizi AWS :
   a. Per proteggere i carichi di lavoro utilizzando EC2 oECS, implementa un Elastic Load Balancer davanti alle risorse di elaborazione. Elastic Load Balancing fornisce quindi la soluzione per rilevare le istanze in zone non integre e instradare il traffico verso quelle integre.
      i. Getting started with Application Load Balancers
      ii. Getting started with Network Load Balancers
   b. Nel caso di EC2 istanze che eseguono off-the-shelf software commerciale che non supporta il bilanciamento del carico, è possibile ottenere una forma di tolleranza agli errori implementando una metodologia di disaster recovery Multi-AZ.
      i. the section called “REL13-BP02 Utilizzare strategie di ripristino definite per raggiungere gli obiettivi di ripristino”
   c. Per ECS le attività di Amazon, distribuisci il servizio in modo uniforme su tre aree AZs per raggiungere un equilibrio tra disponibilità e costi.
      i. Best practice sulla ECS disponibilità di Amazon | Contenitori
   d. Per Amazon non auroraRDS, puoi scegliere Multi-AZ come opzione di configurazione. In caso di guasto dell'istanza del database principale, Amazon promuove RDS automaticamente un database in standby per ricevere traffico in un'altra zona di disponibilità. Puoi inoltre creare repliche di lettura multi-regione per migliorare la resilienza.
      i. Implementazioni Amazon RDS Multi AZ
      ii. Creazione di una replica di lettura in un altro Regione AWS
3. Ecco alcuni esempi di funzionalità multiregionali fornite dai AWS servizi:
   a. Per i carichi di lavoro Amazon S3 in cui la disponibilità multi-AZ è fornita automaticamente dal servizio, considera i punti di accesso multi-regione se è necessaria un'implementazione di questo tipo.
      i. Multi-Region Access Points in Amazon S3
   b. Per le tabelle DynamoDB in cui la disponibilità multi-AZ è fornita automaticamente dal servizio, è possibile convertire facilmente le tabelle esistenti in tabelle globali per sfruttare più regioni.
      i. Convert Your Single-Region Amazon DynamoDB Tables to Global Tables
   c. Se il tuo carico di lavoro è gestito da Application Load Balancer o Network Load Balancer, utilizza AWS Global Accelerator per migliorare la disponibilità dell'applicazione indirizzando il traffico verso più aree che contengono endpoint integri.

   i. Endpoint per acceleratori standard in Global Accelerator - Global Accelerator (amazon.com) AWSAWS

d. Per le applicazioni che sfruttano i vantaggi AWS EventBridge, prendi in considerazione gli autobus interregionali per inoltrare gli eventi ad altre regioni selezionate.

   i. Invio e ricezione di EventBridge eventi Amazon tra Regioni AWS

e. Per i database Amazon Aurora, considera i database Aurora globali, che si estendono su più regioni AWS . I cluster esistenti possono essere modificati per aggiungere anche nuove regioni.

   i. Getting started with Amazon Aurora global databases

f. Se il tuo carico di lavoro include chiavi di crittografia AWS Key Management Service (AWS KMS), valuta se le chiavi multiregionali sono appropriate per la tua applicazione.

   i. Chiavi multiregionali in AWS KMS

g. Per altre funzionalità AWS del servizio, consulta questa serie di blog sulla creazione di un'applicazione multiregionale con AWS servizi

Livello di impegno per il piano di implementazione: da moderato a elevato

Risorse

Documenti correlati:

- Creazione di un'applicazione multiregionale con serie Services AWS
- Architettura di disaster recovery (DR) attiva AWS, parte IV: attiva/attiva su più siti
- Infrastruttura globale di AWS
- Zone locali AWS FAQ
- Architettura di disaster recovery (DR) su AWS, parte I: Strategie per il ripristino nel cloud
- Il ripristino di emergenza è diverso nel cloud
- Global Tables: Multi-Region Replication with DynamoDB

Video correlati:

- AWS re:Invent 2018: modelli di architettura per applicazioni Active-Active in più regioni (09-R2) ARC2
- Auth0: Multi-Region High-Availability Architecture that Scales to 1.5B+ Logins a Month with automated failover

Esempi correlati:

- Architettura di disaster recovery (DR) su, parte I: Strategie per il ripristino nel AWS cloud
- DTCCraggiunge una resilienza che va ben oltre ciò che possono fare in sede
- Expedia Group utilizza un'architettura multiregionale e multi-zona di disponibilità con un DNS servizio proprietario per aggiungere resilienza alle applicazioni
- Uber: Disaster Recovery for Multi-Region Kafka
- Netflix: Active-Active for Multi-Regional Resilience
- How we build Data Residency for Atlassian Cloud
- Intuit funziona in due regioni TurboTax

REL10-BP03 Ripristino automatico dei componenti vincolati a un'unica posizione

Se i componenti del carico di lavoro possono essere eseguiti in una sola zona di disponibilità o in un data center on-premises, devi rendere possibile la ricostruzione completa del carico di lavoro in base agli obiettivi di ripristino definiti.

Livello di rischio associato se questa best practice non fosse adottata: medio

Guida all'implementazione

Se, a causa di vincoli tecnologici, non è possibile seguire le linee guida per distribuire il carico di lavoro in più posizioni, è necessario implementare un percorso alternativo mirato alla resilienza. È necessario automatizzare la possibilità di ricreare l'infrastruttura necessaria, ridistribuire le applicazioni e ricreare i dati necessari per questi casi.

Ad esempio, Amazon EMR avvia tutti i nodi per un determinato cluster nella stessa zona di disponibilità perché l'esecuzione di un cluster nella stessa zona migliora le prestazioni dei flussi di lavoro in quanto fornisce una velocità di accesso ai dati più elevata. Se questo componente è necessario per la resilienza del carico di lavoro, è necessario disporre di un modo per implementare nuovamente il cluster e i relativi dati. Inoltre, per AmazonEMR, dovresti fornire la ridondanza in modi diversi dall'utilizzo di Multi-AZ. Puoi effettuare il provisioning di più nodi. Utilizzando EMRFile System (EMRFS), i dati in ingresso EMR possono essere archiviati in Amazon S3, che a sua volta può essere replicato su più zone di disponibilità oppure. Regioni AWS

Analogamente, per Amazon Redshift, per impostazione predefinita effettua il provisioning del cluster in una zona di disponibilità selezionata casualmente all' Regione AWS interno di quella selezionata. Viene effettuato il provisioning di tutti i nodi del cluster nella stessa zona.

Per i carichi di lavoro basati su server con stato distribuiti in un data center locale, puoi utilizzarli per proteggere i tuoi carichi di lavoro in. AWS Elastic Disaster Recovery AWS Se sei già ospitato in AWS, puoi utilizzare Elastic Disaster Recovery per proteggere il tuo carico di lavoro in una zona o regione di disponibilità alternativa. Elastic Disaster Recovery sfrutta la replica a livello di blocco continua in un'area di gestione temporanea leggera per fornire il ripristino rapido e affidabile di applicazioni on-premises e basate sul cloud.

Passaggi dell'implementazione

1. Implementa l'autoriparazione. Implementa istanze o container utilizzando, quando possibile, il dimensionamento automatico. Se non puoi utilizzare il ridimensionamento automatico, utilizza il ripristino automatico per EC2 le istanze o implementa l'automazione con riparazione automatica basata su Amazon EC2 o sugli eventi del ciclo di vita dei ECS container.
   - Utilizza i gruppi di Amazon EC2 Auto Scaling per istanze e carichi di lavoro di container che non richiedono un indirizzo IP di singola istanza, un indirizzo IP privato, un indirizzo IP elastico e metadati dell'istanza.
     - È possibile usare i dati utente del modello di avvio per implementare l'automazione per la riparazione automatica della maggior parte dei carichi di lavoro.
   - Utilizza il ripristino automatico delle EC2 istanze Amazon per carichi di lavoro che richiedono un indirizzo ID di istanza singola, un indirizzo IP privato, un indirizzo IP elastico e metadati dell'istanza.
     - Automatic Recovery invierà avvisi sullo stato del ripristino a un SNS argomento non appena viene rilevato l'errore dell'istanza.
   - Utilizza gli eventi del ciclo di vita delle EC2 istanze Amazon o gli ECS eventi Amazon per automatizzare la riparazione automatica laddove non è possibile utilizzare la scalabilità o EC2 il ripristino automatici.
     - Utilizza gli eventi per richiamare l'automazione che riparerà il tuo componente secondo la logica di processo richiesta.
   - Utilizza AWS Elastic Disaster Recovery per proteggere i carichi di lavoro stateful limitati a una singola posizione.

Risorse

Documenti correlati:

- ECSEventi Amazon

- Ganci per il ciclo di vita di Amazon EC2 Auto Scaling
- Recupero di un'istanza
- Ridimensionamento automatico del servizio
- Che cos'è Amazon EC2 Auto Scaling?
- AWS Elastic Disaster Recovery

REL10-BP04 Usa architetture a paratia per limitare l'ambito dell'impatto

Implementa architetture a scomparti (note anche come architetture basate su celle) per limitare l'effetto di un guasto all'interno di un carico di lavoro a un numero ridotto di componenti.

Risultato desiderato: un'architettura basata su celle utilizza più istanze isolate di un carico di lavoro, ciascuna delle quali è nota come cella. Ogni cella è indipendente, non condivide lo stato con altre celle e gestisce un sottoinsieme delle richieste complessive del carico di lavoro. Questo approccio riduce il possibile impatto di un errore, ad esempio un aggiornamento software non valido, a una singola cella e alle richieste elaborate. Se un carico di lavoro usa 10 celle per gestire 100 richieste e si verifica un errore, il 90% delle richieste complessive non sarà interessato dall'errore.

Anti-pattern comuni:

- Aumento illimitato delle celle.
- Applicazione di aggiornamenti o implementazioni del codice in tutte le celle contemporaneamente.
- Condivisione dello stato dei componenti tra celle (con l'eccezione del livello di instradamento).
- Aggiunta di logica di business o instradamento complessa al livello di instradamento.
- Le interazioni tra celle non sono ridotte al minimo.

Vantaggi dell'adozione di questa best practice: limitazione alla cella stessa di molti tipi comuni di errori, a garanzia di un ulteriore isolamento dei guasti, grazie alle architetture basate su celle. Questi limiti relativi agli errori possono garantire resilienza in caso di determinati tipi di errori, altrimenti difficili da contenere, come implementazioni di codice non riuscite o richieste danneggiate o che richiamano una modalità di errore specifica (nota anche come richieste poison pill).

Livello di rischio associato se questa best practice non fosse adottata: elevato

Guida all'implementazione

Su una nave gli scomparti permettono di limitare la falla di uno scafo a una sola sezione dello scafo. In sistemi complessi, questo modello viene spesso replicato per consentire l'isolamento degli errori. Le limitazioni per l'isolamento degli errori riducono l'effetto di un errore all'interno di un carico di lavoro a un numero limitato di componenti. I componenti al di fuori della barriera non subiscono gli effetti del guasto. Utilizzando più barriere per l'isolamento dei guasti, puoi limitare l'impatto sul carico di lavoro. No AWS, i clienti possono utilizzare più zone e regioni di disponibilità per fornire l'isolamento dai guasti, ma il concetto di isolamento dei guasti può essere esteso anche all'architettura del carico di lavoro.

Il carico di lavoro complessivo viene partizionato in celle tramite una chiave di partizione. Questa chiave deve essere allineata alla granularità del servizio o al modo naturale in cui il carico di lavoro del servizio può essere suddiviso con interazioni minime tra celle. Esempi di chiavi di partizione sono l'ID cliente, l'ID della risorsa o qualsiasi altro parametro facilmente accessibile nella maggior parte delle API chiamate. Un livello di instradamento alle celle distribuisce le richieste a singole celle in base alla chiave di partizione e presenta un unico endpoint ai client.

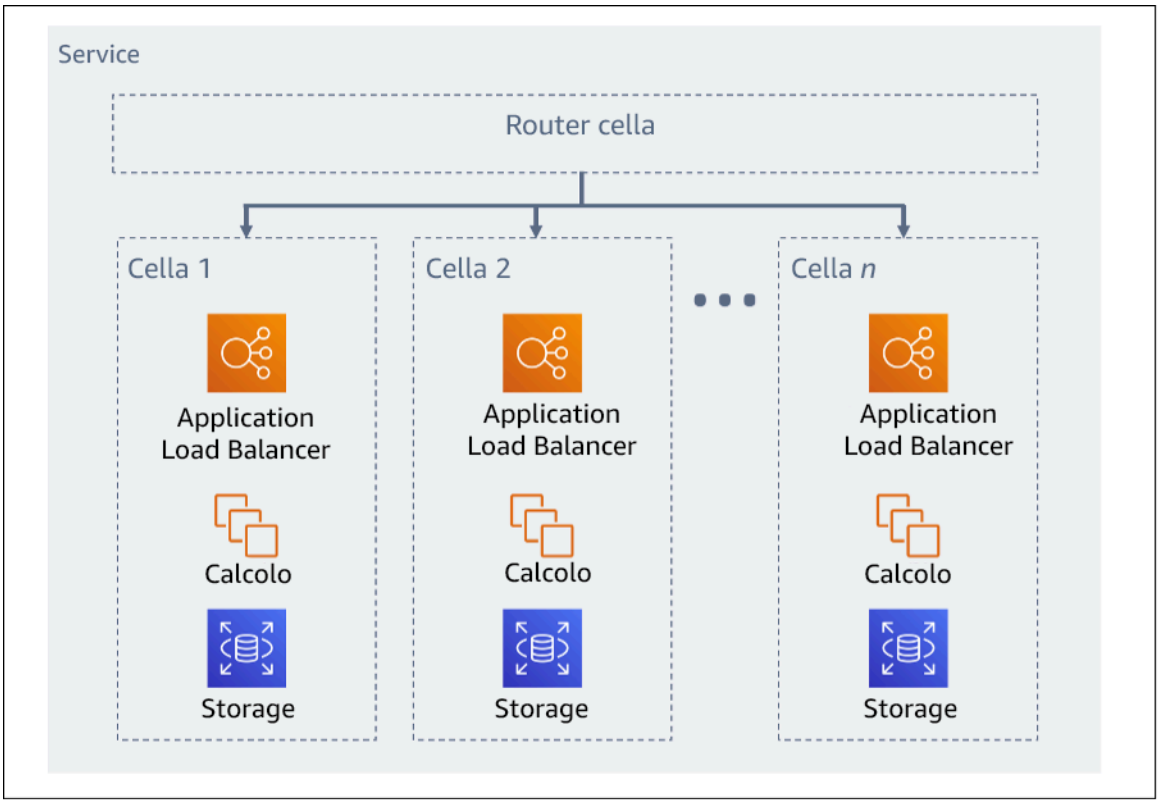


Figura 11: architettura basata su celle

Passaggi dell'implementazione

Nel progettare un'architettura basata su celle, devi tenere conto di diversi aspetti della progettazione:

1. Chiave di partizione: presta particolare attenzione alla scelta della chiave di partizione.
   - Questa deve essere allineata alla granularità del servizio o al modo naturale in cui il carico di lavoro del servizio può essere suddiviso con interazioni minime tra celle. Alcuni esempi sono `customer ID` o `resource ID`.
   - La chiave di partizione deve essere disponibile in tutte le richieste, direttamente o in modo da poter essere facilmente dedotta in modo deterministico da altri parametri.
2. Mappatura persistente delle celle: i servizi a monte devono interagire solo con una singola cella per l'intero ciclo di vita delle risorse correlate.
   - A seconda del carico di lavoro, può essere necessaria una strategia di migrazione delle celle per la migrazione dei dati da una cella a un'altra. Un possibile scenario in cui è necessaria la migrazione delle celle è quando una risorsa o un utente specifico nel carico di lavoro diventa troppo grande e richiede una cella dedicata.
   - Le celle non devono condividere lo stato o i componenti.
   - Di conseguenza, l'interazione tra celle deve essere evitata o mantenuta al minimo, in quanto le interazioni creano dipendenze tra le celle e riducono quindi i vantaggi forniti dall'isolamento degli errori.
3. Livello di instradamento: il livello di instradamento è un componente condiviso tra celle, pertanto non può basarsi sulla stessa strategia di compartimentazione delle celle.
   - È consigliabile che il livello di instradamento distribuisca richieste a singole celle usando un algoritmo di mappatura delle partizioni efficiente in termini di risorse di calcolo, ad esempio combinando funzioni hash crittografiche e aritmetica modulare per mappare le chiavi di partizione alle celle.
   - Per evitare l'impatto su più celle, il livello di instradamento deve restare il più semplice e orizzontalmente scalabile possibile, evitando logica di business complessa in questo livello. Questo approccio offre il vantaggio aggiuntivo di semplificare la comprensione del suo comportamento previsto in ogni momento, permettendo test esaustivi. Come illustrato da Colm MacCárthaigh in Reliability, constant work, and a good cup of coffee, progettazioni semplici e schemi di lavoro costanti si traducono in sistemi affidabili e nella riduzione dell'antifragilità.
4. Dimensione delle celle: le celle devono avere una dimensione massima che non deve essere superata
   - La dimensione massima va identificata attraverso l'esecuzione di test completi, fino a raggiungere i punti di rottura e definire i margini operativi. Per ulteriori informazioni su come implementare procedure di test, consulta REL07-BP04 Load Testa il tuo carico di lavoro.

- L'aumento del carico di lavoro complessivo deve essere gestito tramite l'aggiunta di celle, in modo da poterlo dimensionare in base al crescere della domanda.

5. Strategie multi-AZ o multi-regione: si consiglia di utilizzare più livelli di resilienza per proteggersi da diversi domini di errore.
   - Per la resilienza, devi utilizzare un approccio che costruisca livelli di difesa. Un livello protegge da interruzioni più piccole e più comuni creando un'architettura ad alta disponibilità che utilizza più elementi. AZs Un altro livello di difesa è destinato a proteggere da eventi rari come disastri naturali diffusi e interruzioni a livello regionale. Questo secondo livello prevede l'architettura dell'applicazione in modo che si estenda su più fronti. Regioni AWS L'implementazione di una strategia multi-regione per il tuo carico di lavoro aiuta a proteggerlo da disastri naturali diffusi che colpiscono un'ampia regione geografica di un paese o da guasti tecnici di portata regionale. Tieni presente che l'implementazione di un'architettura multi-regione può essere molto complessa e di solito non è necessaria per la maggior parte dei carichi di lavoro. Per ulteriori dettagli, consulta REL10-BP02 Seleziona le posizioni appropriate per l'implementazione in più sedi.
6. Implementazione del codice: è preferibile una strategia di implementazione del codice scaglionata rispetto all'implementazione simultanea di modifiche al codice in tutte le celle.
   - In questo modo, è possibile ridurre al minimo eventuali errori in più celle a causa di un'implementazione non corretta o dell'errore umano. Per ulteriori informazioni, consulta Automatizzazione di distribuzioni pratiche e sicure.

Risorse

Best practice correlate:

- REL07-BP04 Load Testa il tuo carico di lavoro
- REL10-BP02 Seleziona le posizioni appropriate per l'implementazione in più sedi

Documenti correlati:

- Reliability, constant work, and a good cup of coffee
- AWS e compartimentazione
- Isolamento del carico di lavoro utilizzando lo sharding casuale
- Automatizzazione di distribuzioni pratiche e sicure

Video correlati:

- AWS re:Invent 2018: Chiudere i circuiti e aprire le menti: come assumere il controllo di sistemi, grandi e piccoli
- AWS re:Invent 2018: Come AWS ridurre al minimo il raggio di esplosione dei guasti () ARC338
- Shuffle-sharding: AWS re:Invent 2019: presentazione della libreria Amazon Builders () DOP328
- AWS Summit ANZ 2021 - Tutto fallisce, in continuazione: progettare per la resilienza

Esempi correlati:

- Well-Architected Lab: isolamento degli errori con il partizionamento casuale

REL11. Come si progetta il carico di lavoro affinché resista ai guasti dei componenti?

I carichi di lavoro che richiedono un'elevata disponibilità e un basso tempo medio di ripristino (MTTR) devono essere progettati in modo da garantire la resilienza.

Best practice

- REL11-BP01 Monitora tutti i componenti del carico di lavoro per rilevare i guasti
- REL11-BP02 Fallimento verso risorse sane
- REL11-BP03 Automatizza la guarigione su tutti i livelli
- REL11-BP04 Affidati al piano dati e non al piano di controllo durante il ripristino
- REL11-BP05 Usa la stabilità statica per prevenire il comportamento bimodale
- REL11-BP06 Invia notifiche quando gli eventi influiscono sulla disponibilità
- REL11-BP07 Progetta il tuo prodotto per soddisfare gli obiettivi di disponibilità e gli accordi sui livelli di servizio di uptime () SLAs

REL11-BP01 Monitora tutti i componenti del carico di lavoro per rilevare i guasti

Monitora costantemente lo stato del carico di lavoro, in modo che tu e i tuoi sistemi automatizzati siate consapevoli di errori o guasti non appena si verificano. Monitora gli indicatori chiave di performance (KPIs) in base al valore aziendale.

Tutti i meccanismi di ripristino e correzione devono essere in grado di rilevare rapidamente i problemi. I guasti tecnici devono essere rilevati prima in modo che possano essere risolti. Tuttavia,

la disponibilità si basa sulla capacità del carico di lavoro di fornire valore aziendale, pertanto gli indicatori chiave di performance (KPIs) che misurano questo fattore devono far parte della strategia di rilevamento e correzione.

Risultato desiderato: i componenti essenziali di un carico di lavoro vengono monitorati in modo indipendente per rilevare guasti e fornire avvisi quando e dove si verificano.

Anti-pattern comuni:

- Non sono stati configurati allarmi, pertanto le interruzioni si verificano senza notifica.
- Gli allarmi esistono, ma a soglie che non forniscono tempo adeguato per reagire.
- Le metriche non vengono raccolte abbastanza spesso per soddisfare l'obiettivo del tempo di ripristino (). RTO
- Solo le interfacce del carico di lavoro rivolte al cliente vengono monitorate attivamente.
- Viene effettuata solo la raccolta di parametri tecnici, senza includere quelli delle funzioni aziendali.
- Non è presente alcun parametro che misuri l'esperienza utente del carico di lavoro.
- Vengono creati troppi monitoraggi.

Vantaggi dell'adozione di questa best practice: eseguire un monitoraggio appropriato a tutti i livelli consente di ridurre i tempi di rilevamento, velocizzando quindi il ripristino.

Livello di rischio associato se questa best practice non fosse adottata: elevato

Guida all'implementazione

Identifica tutti i carichi di lavoro che verranno esaminati per il monitoraggio. Dopo aver identificato tutti i componenti del carico di lavoro da monitorare, devi determinare l'intervallo di monitoraggio. L'intervallo di monitoraggio ha un impatto diretto sulla velocità con cui il ripristino viene avviato, che dipende dal tempo impiegato per rilevare un errore. Il tempo medio di rilevamento (MTTD) è la quantità di tempo che intercorre tra il verificarsi di un guasto e l'inizio delle operazioni di riparazione. L'elenco dei servizi deve essere ampio e completo.

Il monitoraggio deve includere tutti i livelli dello stack applicativo, come applicazione, piattaforma, infrastruttura e rete.

La strategia di monitoraggio deve tenere in considerazione l'impatto dei guasti nell'area grigia. Per ulteriori informazioni sui guasti nell'area grigia, consulta Gray failures nel whitepaper Advanced Multi-AZ Resilience Patterns.

Passaggi dell'implementazione

- L'intervallo di monitoraggio dipende dalla velocità con cui è necessario ripristinare Il tempo di ripristino dipende dal tempo necessario per il ripristino, pertanto è necessario determinare la frequenza di raccolta tenendo conto di questo tempo e dell'obiettivo del tempo di ripristino (RTO).
- Configura il monitoraggio dettagliato per componenti e servizi gestiti.
  - Determina se è necessario un monitoraggio dettagliato EC2 delle istanze e dell'Auto Scaling. Il monitoraggio dettagliato fornisce metriche a intervalli di un minuto, mentre il monitoraggio predefinito fornisce metriche a intervalli di cinque minuti.
  - Determina se RDS è necessario un monitoraggio avanzato per. Il monitoraggio avanzato utilizza un agente sulle RDS istanze per ottenere informazioni utili su diversi processi o thread.
  - Determina i requisiti di monitoraggio dei componenti serverless critici per Lambda API, Gateway, AmazonEKS, ECSAmazon e tutti i tipi di sistemi di bilanciamento del carico.
  - Determina i requisiti di monitoraggio dei componenti di storage per Amazon S3, AmazonEFS, FSx Amazon e Amazon. EBS
- Crea metriche personalizzate per misurare gli indicatori chiave di performance aziendali ()KPIs. I carichi di lavoro implementano funzioni aziendali chiave, che dovrebbero essere utilizzate in quanto KPIs aiutano a identificare quando si verifica un problema indiretto.
- Monitora la presenza di errori nell'esperienza utente tramite le canary degli utenti Il test sintetico delle transazioni (noto anche come "test canary", ma da non confondere con le distribuzioni canary) in grado di eseguire e simulare il comportamento dei clienti è uno dei processi di test più importanti. Esegui questi test costantemente sugli endpoint del carico di lavoro da diverse posizioni remote.
- Crea parametri personalizzati che monitorino l'esperienza dell'utente. Dotare l'esperienza del cliente di strumenti consente di determinare quando essa peggiora.
- Imposta gli allarmi per rilevare quando una qualsiasi parte del carico di lavoro non funziona correttamente e per indicare quando effettuare il dimensionamento automatico delle risorse. Gli allarmi possono essere visualizzati visivamente sulle dashboard, inviare avvisi tramite Amazon SNS o e-mail e utilizzare Auto Scaling per aumentare o ridurre le risorse del carico di lavoro.
- Crea pannelli di controllo per visualizzare i parametri. Utilizza i pannelli di controllo per visualizzare tendenze, valori anomali e altri indicatori di potenziali problemi, oppure per fornire un'indicazione dei problemi che potresti voler approfondire.
- Crea il monitoraggio del tracciamento distribuito per i tuoi servizi. Con il monitoraggio distribuito puoi comprendere le prestazioni della tua applicazione e dei relativi servizi sottostanti per identificare e risolvere la causa ultima di problemi ed errori riguardanti le prestazioni.

- Crea dashboard di sistemi di monitoraggio (utilizzando CloudWatcho X-Ray) e raccolta dati in una regione e in un account separati.
- Crea un'integrazione per il monitoraggio di Amazon Health Aware per consentire il monitoraggio della visibilità AWS delle risorse che potrebbero presentare un deterioramento. Per i carichi di lavoro aziendali essenziali, questa soluzione fornisce l'accesso ad avvisi proattivi e in tempo reale per i servizi. AWS

Risorse

Best practice correlate:

- Definizione di disponibilità
- REL11-BP06 Invia notifiche quando gli eventi influiscono sulla disponibilità

Documenti correlati:

- Amazon CloudWatch Synthetics ti consente di creare canari utente
- Abilitare o disabilitare il monitoraggio dettagliato della propria istanza
- Monitoraggio avanzato
- Monitoraggio dei gruppi e delle istanze di Auto Scaling tramite Amazon CloudWatch
- Publishing Custom Metrics
- Utilizzo di Amazon CloudWatch Alarms
- Utilizzo dei pannelli di controllo CloudWatch
- Utilizzo di dashboard Cross Region Cross Account CloudWatch
- Uso del tracciamento X-Ray tra più regioni e account
- Understanding availability
- Implementazione di Amazon Health Aware (AHA)

Video correlati:

- Mitigating gray failures

Esempi correlati:

- Well-Architected Lab (Livello 300): Implementing Health Checks and Managing Dependencies to Improve Reliability
- One Observability Workshop: Explore X-Ray

Strumenti correlati:

- CloudWatch
- CloudWatch X-Ray

REL11-BP02 Fallimento verso risorse sane

Se si verifica un errore in una risorsa, le risorse integre dovrebbero continuare a soddisfare le richieste. In caso di problemi di ubicazione (come la zona di disponibilità o Regione AWS), assicurati di disporre di sistemi che consentano il failover su risorse integre in aree prive di problemi.

Durante la progettazione di un servizio, distribuisci il carico tra risorse, zone di disponibilità o regioni. In questo modo, il guasto o la compromissione di una singola risorsa può essere mitigato spostando il traffico sulle risorse integre rimanenti. Considera come vengono rilevati e indirizzati i servizi in caso di guasto.

Progetta i tuoi servizi tenendo a mente il recupero dai guasti. In AWS, progettiamo servizi per ridurre al minimo i tempi di ripristino in caso di guasti e l'impatto sui dati. I nostri servizi utilizzano principalmente archivi di dati che riconoscono le richieste solo dopo che queste sono state archiviate in modo duraturo su più repliche in una regione. Sono costruiti con il criterio dell'isolamento basato sulle celle ed utilizzano l'isolamento dei guasti fornito dalle zone di disponibilità. Facciamo ampio uso dell'automazione nelle nostre procedure operative. Ottimizziamo inoltre la nostra replace-and-restart funzionalità per ripristinare rapidamente le interruzioni.

I modelli e i progetti che consentono il failover variano a seconda dei servizi della AWS . Molti servizi gestiti AWS nativi sono nativamente più zone di disponibilità (come Lambda API o Gateway). Altri AWS servizi (come EC2 andEKS) richiedono una progettazione basata su best practice specifiche per supportare il failover delle risorse o lo storage dei dati in tutte le aree. AZs

Il monitoraggio deve essere impostato per verificare che la risorsa di failover sia integra, tenere traccia dell'avanzamento del failover delle risorse e monitorare il ripristino dei processi aziendali.

Risultato desiderato: i sistemi sono in grado di utilizzare automaticamente o manualmente nuove risorse per il ripristino dopo un evento di deterioramento.

Anti-pattern comuni:

- La pianificazione degli errori non fa parte della fase di pianificazione e progettazione.
- RTOe non RPO sono stabiliti.
- Monitoraggio insufficiente per rilevare risorse difettose.
- Isolamento adeguato dei domini di errore.
- Il failover multi-regione non è considerato.
- Il rilevamento dei guasti è troppo sensibile o aggressivo quando si decide di eseguire il failover.
- Non è possibile testare o convalidare il progetto di failover.
- Esecuzione dell'automazione del risanamento automatico, ma senza la notifica della necessità di una correzione.
- Mancanza di un periodo di mitigazione per evitare che l'errore si ripresenti troppo presto.

Vantaggi dell'adozione di questa best practice: è possibile creare sistemi più resilienti che garantiscano l'affidabilità in caso di guasti eseguendo prima un deterioramento lento e poi un ripristino rapido.

Livello di rischio associato se questa best practice non fosse adottata: elevato

Guida all'implementazione

AWS i servizi, come Elastic Load Balancing e Amazon Auto EC2 Scaling, aiutano a distribuire il carico tra risorse e zone di disponibilità. Pertanto, il guasto di una singola risorsa (come un'EC2istanza) o il danneggiamento di una zona di disponibilità possono essere mitigati spostando il traffico verso le risorse rimanenti integre.

Per i carichi di lavoro multi-regione, i progetti sono più complicati. Ad esempio, le repliche di lettura tra aree geografiche consentono di distribuire i dati su più aree. Regioni AWS Tuttavia, il failover è ancora necessario per promuovere la replica di lettura a principale e quindi indirizzare il traffico verso il nuovo endpoint. Amazon Route 53, Amazon Application Recovery Controller (ARC) CloudFront, Amazon e AWS Global Accelerator possono aiutare a instradare il traffico Regioni AWS.

AWS i servizi, come Amazon S3, LambdaAPI, Gateway, Amazon, Amazon, SQS Amazon, SNS Amazon SES Pinpoint, Amazon AWS Certificate Manager o EventBridge Amazon DynamoDBECR, vengono distribuiti automaticamente in più zone di disponibilità da. AWS In caso di guasto, questi AWS servizi indirizzano automaticamente il traffico verso luoghi integri. I dati sono archiviati in modo ridondante in più zone di disponibilità e rimangono disponibili.

Per AmazonRDS, Amazon Aurora, Amazon Redshift, Amazon o EKS ECS Amazon, Multi-AZ è un'opzione di configurazione. AWS può indirizzare il traffico verso l'istanza integra se viene avviato il failover. Questa azione di failover può essere intrapresa dal cliente AWS o in base alle sue esigenze

Per EC2 le istanze Amazon, Amazon Redshift, Amazon Tasks o ECS EKS Amazon Pods, sei tu a scegliere in quali zone di disponibilità eseguire la distribuzione. Per alcuni progetti, Elastic Load Balancing fornisce la soluzione per rilevare le istanze in zone non integre e instradare il traffico verso quelle integre. Elastic Load Balancing può inoltre instradare il traffico verso componenti nel tuo data center on-premises.

Per il failover del traffico multiregionale, il reindirizzamento può sfruttare Amazon Route 53, Amazon Application Recovery Controller AWS Global Accelerator e Route 53 DNS Private VPCs per CloudFront o fornire un modo per definire domini Internet e assegnare politiche di routing, compresi i controlli di integrità, per instradare il traffico verso regioni integre. AWS Global Accelerator fornisce indirizzi IP statici che fungono da punto di accesso fisso all'applicazione, quindi indirizzano verso gli endpoint Regioni AWS di propria scelta, utilizzando la rete AWS globale anziché Internet per prestazioni e affidabilità migliori.

Passaggi dell'implementazione

- Crea progetti di failover per tutte le applicazioni e i servizi appropriati. Isola ogni componente dell'architettura e crea progetti di failover che RTO soddisfino le esigenze di RPO ciascun componente.
- Configura ambienti inferiori (come sviluppo o test) con tutti i servizi necessari per disporre di un piano di failover. Implementa le soluzioni utilizzando il modello infrastructure as code (IaC) per garantire la ripetibilità.
- Configura un sito di ripristino, ad esempio una seconda regione, per implementare e testare i progetti di failover. Se necessario, le risorse per i test possono essere configurate temporaneamente per limitare i costi aggiuntivi.
- Determina da quali piani di failover vengono automatizzati AWS, quali possono essere automatizzati mediante un DevOps processo e quali possono essere manuali. Documenta e misura ogni servizio RTO eRPO.
- Crea un playbook per il failover e includi tutti i passaggi necessari per eseguire il failover di ogni risorsa, applicazione e servizio.
- Crea un playbook di failback e includi tutti i passaggi per eseguire il failback (con tempistiche) di ogni risorsa, applicazione e servizio.

- Crea un piano per avviare e testare il playbook. Usa simulazioni e test del caos per testare i passaggi e l'automazione del playbook.
- In caso di problemi di ubicazione (ad esempio nella zona di disponibilità o Regione AWS), assicurati di disporre di sistemi che consentano il failover su risorse integre in luoghi integri. Verifica la quota, i livelli di dimensionamento automatico e le risorse in esecuzione prima dei test di failover.

Risorse

Best practice Well-Architected correlate:

- REL13- Piano per il DR
- REL10 - Utilizza l'isolamento dai guasti per proteggere il carico di lavoro

Documenti correlati:

- Impostazione RTO e RPO obiettivi
- Failover utilizzando il routing ponderato Route 53
- Ripristino di emergenza con Amazon Application Recovery Controller
- EC2con scalabilità automatica
- EC2Implementazioni: Multi-AZ
- ECSImplementazioni: Multi-AZ
- Scambia il traffico utilizzando Amazon Application Recovery Controller
- Lambda with an Application Load Balancer and Failover
- ACMReplica e failover
- Parameter Store Replication and Failover
- ECRreplica e failover tra regioni
- Secrets manager cross region replication configuration
- Abilita la replica interregionale e il failover EFS
- EFSReplica e failover tra regioni
- Networking Failover
- Failover degli endpoint S3 utilizzando MRAP
- Crea una replica tra regioni per S3
- Guida per il failover interregionale e l'attivazione di Graceful Failback AWS

- Failover using multi-region global accelerator
- Failover con DRS
- Creating Disaster Recovery Mechanisms Using Amazon Route 53

Esempi correlati:

- Disaster Recovery attivo AWS
- Elastic Disaster Recovery attivo AWS

REL11-BP03 Automatizza la guarigione su tutti i livelli

Al rilevamento di un guasto, utilizza funzionalità automatizzate per eseguire azioni da correggere. I guasti possono essere riparati automaticamente tramite meccanismi di servizio interni oppure riavviando o rimuovendo le risorse tramite azioni correttive.

Per applicazioni gestite dal cliente e per il ripristino tra regioni, è possibile attingere a modelli di ripristino e processi di riparazione automatizzati dalle best practice esistenti.

La possibilità di riavviare o rimuovere una risorsa è uno strumento importante per risolvere i guasti. Una best practice consiste nel rendere i servizi stateless, ove possibile. In questo modo si evita la perdita di dati o di disponibilità durante il riavvio della risorsa. Nel cloud è possibile, e in genere si dovrebbe, sostituire l'intera risorsa (ad esempio, un'istanza di calcolo o una funzione serverless) come parte del riavvio. Il riavvio stesso è un modo semplice e affidabile per eseguire il ripristino in caso di guasto. Molti tipi diversi di guasto si verificano nei carichi di lavoro. Possono verificarsi guasti a livello di hardware, software, comunicazione e operazioni.

Il riavvio o i nuovi tentativi come pratiche risolutive si applicano anche alle richieste di rete. Adotta lo stesso approccio di ripristino sia a un timeout di rete sia a un guasto di dipendenza in cui la dipendenza restituisce un guasto. Entrambi gli eventi hanno un effetto simile sul sistema, quindi piuttosto che tentare di trasformare entrambi gli eventi in un caso speciale, adotta una strategia analoga di nuovi tentativi limitati con un un jitter e un backoff esponenziali. La capacità di riavvio è un meccanismo di ripristino presente nelle architetture di cluster ROC (Recovery-oriented computing) e ad alta disponibilità.

Risultato desiderato: vengono eseguite azioni automatiche di risoluzione a seguito del rilevamento di un errore.

Anti-pattern comuni:

- Provisioning di risorse senza dimensionamento automatico.
- Implementazione individuale di applicazioni in istanze/container.
- Implementazione di applicazioni che non possono essere distribuite in più posizioni senza utilizzare il ripristino automatico.
- Riparazione manuale delle applicazioni che il dimensionamento e il ripristino automatici non sono stati in grado di riparare.
- Nessuna automazione dei database di failover.
- Mancanza di metodi automatizzati per reinstradare il traffico verso nuovi endpoint.
- Nessuna replica dell'archiviazione.

Vantaggi dell'adozione di questa best practice: la riparazione automatica può ridurre il tempo medio di ripristino e migliorare la disponibilità.

Livello di rischio associato se questa best practice non fosse adottata: elevato

Guida all'implementazione

I progetti per Amazon EKS o altri servizi Kubernetes devono includere set di repliche o stateful set minimi e massimi e il dimensionamento minimo di cluster e gruppi di nodi. Questi meccanismi forniscono una quantità minima di risorse di elaborazione continuamente disponibili mentre riparano automaticamente eventuali guasti utilizzando il piano di controllo (control-plane) Kubernetes.

I modelli di progettazione a cui si accede tramite un bilanciatore del carico che utilizza cluster di calcolo dovrebbero sfruttare i gruppi Auto Scaling. Elastic Load Balancing (ELB) distribuisce automaticamente il traffico delle applicazioni in entrata su più destinazioni e appliance virtuali in una o più zone di disponibilità (). AZs

I progetti basati su cluster computing che non utilizzano il bilanciamento del carico devono avere dimensioni progettate per la perdita di almeno un nodo. Ciò consentirà al servizio di rimanere in esecuzione con una capacità potenzialmente ridotta durante il ripristino di un nuovo nodo. Servizi di esempio sono Mongo, DynamoDB Accelerator, Amazon Redshift, Amazon, Cassandra, EMR MSK KafkaEC2, -, Couchbase e Amazon Service. ELK OpenSearch Molti di questi servizi possono essere progettati con funzionalità di riparazione automatica aggiuntive. Alcune tecnologie di cluster devono generare un avviso in caso di perdita di un nodo attivando un flusso di lavoro automatico o manuale per creare un nuovo nodo. Questo flusso di lavoro può essere automatizzato per risolvere rapidamente i problemi. AWS Systems Manager

Amazon EventBridge può essere usato per monitorare e filtrare eventi come CloudWatch allarmi o cambiamenti di stato in altri AWS servizi. In base alle informazioni sugli eventi, può quindi richiamare AWS Lambda Systems Manager Automation o altri obiettivi per eseguire una logica di correzione personalizzata sul carico di lavoro. Amazon EC2 Auto Scaling può essere configurato per verificare lo stato delle EC2 istanze. Se l'istanza si trova in uno stato diverso da quello in esecuzione o se lo stato del sistema è compromesso, Amazon EC2 Auto Scaling considera l'istanza non integra e avvia un'istanza sostitutiva. Per le sostituzioni su larga scala (ad esempio la perdita di un'intera zona di disponibilità), è preferibile adottare la stabilità statica per ottenere un'elevata disponibilità.

Passaggi dell'implementazione

- Utilizza i gruppi Auto Scaling per implementare livelli in un carico di lavoro. Auto Scaling è in grado di eseguire il risanamento automatico sulle applicazioni stateless e aggiungere o rimuovere capacità.
- Per le istanze di calcolo menzionate in precedenza, utilizza il bilanciamento del carico e scegli il tipo di bilanciatore del carico adeguato.
- Prendi in considerazione la possibilità di guarire per AmazonRDS. Utilizzando le istanze di standby, puoi configurare il failover automatico sulle stesse. Per Amazon Read RDS Replica, è necessario un flusso di lavoro automatizzato per rendere primaria una replica di lettura.
- Implementa il ripristino automatico su EC2 istanze in cui sono distribuite applicazioni che non possono essere distribuite in più posizioni e può tollerare il riavvio in caso di guasto. Il ripristino automatico può essere utilizzato per sostituire l'hardware guasto e riavviare l'istanza quando l'applicazione non è in grado di essere distribuita in più posizioni. I metadati dell'istanza e gli indirizzi IP associati vengono conservati, nonché EBSi volumi e i punti di montaggio su Amazon Elastic File System o File Systems for Lustre e Windows. Utilizzando AWS OpsWorks, puoi configurare la riparazione automatica delle EC2 istanze a livello di livello.
- Implementa il ripristino automatico utilizzando AWS Step Functions e AWS Lambda quando non è possibile utilizzare il dimensionamento automatico o il ripristino automatico oppure quando il ripristino automatico non riesce. Quando non è possibile utilizzare il ridimensionamento automatico e non è possibile utilizzare il ripristino automatico oppure il ripristino automatico fallisce, è possibile automatizzare la riparazione utilizzando e. AWS Step Functions AWS Lambda
- Amazon EventBridge può essere usato per monitorare e filtrare eventi come CloudWatchallarmi o cambiamenti di stato in altri AWS servizi. In base alle informazioni sugli eventi, può quindi richiamare AWS Lambda (o altre destinazioni) per eseguire una logica di riparazione personalizzata sul tuo carico di lavoro.

Risorse

Best practice correlate:

- Definizione di disponibilità
- REL11-BP01 Monitora tutti i componenti del carico di lavoro per rilevare i guasti

Documenti correlati:

- Funzionamento di AWS Auto Scaling
- EC2Ripristino automatico Amazon
- Amazon Elastic Block Store (AmazonEBS)
- Amazon Elastic File System (AmazonEFS)
- Cos'è Amazon FSx for Lustre?
- Cos'è Amazon FSx per Windows File Server?
- AWS OpsWorks: utilizzo della riparazione automatica per sostituire le istanze con errore
- Che cos'è AWS Step Functions?
- Che cos'è AWS Lambda?
- Che cos'è Amazon EventBridge?
- Utilizzo di Amazon CloudWatch Alarms
- RDSFailover Amazon
- SSM- Systems Manager Automation
- Best practice per architetture resilienti

Video correlati:

- Fornitura e scalabilità automatiche OpenSearch del servizio
- Amazon RDS Failover automatico

Esempi correlati:

- Workshop su Auto Scaling
- Workshop RDS sul failover di Amazon

Strumenti correlati:

- CloudWatch
- CloudWatch X-Ray

REL11-BP04 Affidati al piano dati e non al piano di controllo durante il ripristino

I piani di controllo forniscono le risorse amministrative APIs utilizzate per creare, leggere e descrivere, aggiornare, eliminare ed elencare (CRUDL) le risorse, mentre i piani dati gestiscono il traffico di day-to-day servizio. Durante l'implementazione di risposte di ripristino o mitigazione a eventi che possono influire sulla resilienza, concentrati sull'utilizzo di un numero minimo di operazioni del piano di controllo (control-plane) per ripristinare, ridimensionare, ristabilire, riparare il servizio o eseguirne il failover. Le operazioni del piano dati dovrebbero avere la precedenza su qualsiasi attività durante questi eventi che causano deterioramento.

Ad esempio, le seguenti sono tutte azioni del piano di controllo (control-plane): avvio di una nuova istanza di calcolo, creazione di storage a blocchi e descrizione dei servizi di coda. Quando avvii istanze di calcolo, il piano di controllo (control-plane) deve eseguire diverse attività, come trovare un host fisico con capacità, allocare interfacce di rete, preparare volumi di storage a blocchi locali, generare credenziali e aggiungere regole di sicurezza. I piani di controllo (control-plane) tendono ad avere un'orchestrazione complicata.

Risultato desiderato: quando lo stato di risorsa viene compromesso, il sistema è in grado di ripristinarsi automaticamente o manualmente spostando il traffico da risorse danneggiate a risorse integre.

Anti-pattern comuni:

- Dipendenza dalla modifica DNS dei record per reindirizzare il traffico.
- Dipendenza dalle operazioni di dimensionamento del piano di controllo (control-plane) per sostituire i componenti danneggiati a causa di un provisioning delle risorse insufficiente.
- Affidarsi ad azioni estese, multiservizio e API multicontrollo sul piano per porre rimedio a qualsiasi categoria di deterioramento.

Vantaggi dell'adozione di questa best practice una maggiore percentuale di successo in termini di riparazione automatica può ridurre il tempo medio di ripristino e migliorare la disponibilità del carico di lavoro.

Livello di rischio associato se questa best practice non fosse adottata: medio. Per determinati tipi di degrado del servizio, i piani di controllo (control-plane) sono interessati. La dipendenza dall'uso estensivo del piano di controllo per la riparazione può aumentare il tempo di ripristino () e il tempo medio di ripristino (RTO). MTTR

Guida all'implementazione

Per limitare le azioni del piano dati, esegui una valutazione del servizio determinare le azioni necessarie per ripristinarlo.

Sfrutta Amazon Application Recovery Controller per spostare il DNS traffico. Queste funzionalità monitorano continuamente la capacità dell'applicazione di ripristinarsi in caso di guasti e consentono di controllare il ripristino delle applicazioni su più Regioni AWS zone di disponibilità e in locale.

Le policy di instradamento di Route 53 utilizzano il piano di controllo (control-plane), quindi non fare affidamento su di esso per il ripristino. I piani dati Route 53 rispondono alle DNS domande ed eseguono e valutano i controlli di integrità. Sono distribuiti a livello globale e progettati per un accordo sul livello di servizio con disponibilità del 100% (SLA).

La gestione APIs e le console di Route 53 in cui è possibile creare, aggiornare ed eliminare le risorse Route 53 funzionano su piani di controllo progettati per dare priorità alla forte coerenza e alla durabilità necessarie durante la gestione. DNS A tal fine, i piani di controllo (control-plane) sono situati in un'unica regione: Stati Uniti orientali (Virginia settentrionale). Sebbene entrambi i sistemi siano progettati per essere molto affidabili, i piani di controllo non sono inclusi nel. SLA Possono verificarsi eventi rari in cui la progettazione resiliente del piano dati consente di mantenere la disponibilità mentre i piani di controllo (control-plane) non lo fanno. Per i meccanismi di ripristino di emergenza e failover, utilizzare le funzioni del piano dati per garantire la migliore affidabilità possibile.

Progetta la tua infrastruttura di elaborazione in modo che sia staticamente stabile per evitare di utilizzare il piano di controllo durante un incidente. Ad esempio, se utilizzi EC2 istanze Amazon, evita di effettuare il provisioning di nuove istanze manualmente o di chiedere agli Auto Scaling Groups di aggiungere istanze in risposta. Per ottenere i massimi livelli di resilienza, è necessario fornire una capacità sufficiente nel cluster utilizzato per il failover. Se questa soglia di capacità deve essere limitata, imposta dei limiti sull'intero end-to-end sistema per limitare in modo sicuro il traffico totale che raggiunge il set limitato di risorse.

Per servizi come Amazon DynamoDB, API Amazon Gateway, load balancer e serverless, l'utilizzo di AWS Lambda tali servizi sfrutta il piano dati. Tuttavia, la creazione di nuove funzioni, load balancer, API gateway o tabelle DynamoDB è un'azione del piano di controllo e deve essere completata

prima del degrado come preparazione a un evento e alla ripetizione delle azioni di failover. Per AmazonRDS, le azioni del piano dati consentono l'accesso ai dati.

Per ulteriori informazioni sui piani dati, sui piani di controllo e su come AWS crea servizi per soddisfare gli obiettivi di alta disponibilità, consulta Stabilità statica tramite zone di disponibilità.

Capire quali operazioni sono sul piano dati e quali sul piano di controllo (control-plane).

Passaggi dell'implementazione

Per ogni carico di lavoro che deve essere ripristinato dopo un evento di deterioramento, valuta il runbook di failover, il design ad alta disponibilità, il progetto di riparazione automatica o il piano di ripristino delle risorse HA. Identifica ogni azione che potrebbe essere considerata un'azione del piano di controllo (control-plane).

Prendi in considerazione la possibilità di modificare l'azione di controllo in un'azione del piano dati:

- Auto Scaling (piano di controllo) su risorse EC2 Amazon prescalate (piano dati)
- Scalabilità delle EC2 istanze Amazon (piano di controllo) alla AWS Lambda scalabilità (piano dati)
- Valuta qualsiasi progetto utilizzando Kubernetes e considerando la natura delle azioni del piano di controllo (control-plane). L'aggiunta di pod è un'azione del piano dati in Kubernetes. Le azioni devono limitarsi all'aggiunta di pod e non all'aggiunta di nodi. L'utilizzo di nodi con provisioning eccessivo è il metodo preferibile per limitare le azioni del piano di controllo (control-plane).

Prendi in considerazione approcci alternativi che consentano alle azioni del piano dati di incidere sulla stessa correzione.

- Route 53 Record change (piano di controllo) o Amazon Application Recovery Controller (piano dati)
- Controlli dell'integrità di Route 53 per aggiornamenti più automatizzati

Se il servizio è mission critical, prendi in considerazione alcuni servizi in una regione secondaria per consentire più azioni del piano di controllo (control-plane) e del piano dati in una regione non interessata dal problema.

- Amazon EC2 Auto Scaling o Amazon EKS in una regione principale rispetto ad Amazon Auto EC2 Scaling o EKS Amazon in una regione secondaria e instradamento del traffico verso una regione secondaria (azione del piano di controllo)

- Crea una replica di lettura nella regione secondaria o tenta la stessa azione nella regione principale (azione del piano di controllo (control-plane))

Risorse

Best practice correlate:

- Definizione di disponibilità
- REL11-BP01 Monitora tutti i componenti del carico di lavoro per rilevare i guasti

Documenti correlati:

- APNPartner: partner che possono contribuire all'automazione della tolleranza ai guasti
- Marketplace AWS: prodotti utilizzabili per la tolleranza ai guasti
- Amazon Builders' Library: Avoiding overload in distributed systems by putting the smaller service in control
- Amazon API DynamoDB (piano di controllo e piano dati)
- AWS Lambda Esecuzioni (suddivise in piano di controllo e piano dati)
- AWS Elemental MediaStore Piano dati
- Creazione di applicazioni altamente resilienti utilizzando Amazon Application Recovery Controller, parte 1: stack per regione singola
- Creazione di applicazioni altamente resilienti utilizzando Amazon Application Recovery Controller, parte 2: stack multiregionale
- Creating Disaster Recovery Mechanisms Using Amazon Route 53
- Cos'è Amazon Application Recovery Controller
- Piano di controllo (control-plane) e piano dati di Kubernetes

Video correlati:

- Back to Basics - Using Static Stability
- Creazione di carichi di lavoro resilienti su più siti utilizzando servizi globali AWS

Esempi correlati:

- Presentazione di Amazon Application Recovery Controller

- Amazon Builders' Library: Avoiding overload in distributed systems by putting the smaller service in control
- Creazione di applicazioni altamente resilienti utilizzando Amazon Application Recovery Controller, parte 1: stack per regione singola
- Creazione di applicazioni altamente resilienti utilizzando Amazon Application Recovery Controller, parte 2: stack multiregionale
- Stabilità statica con le zone di disponibilità

Strumenti correlati:

- Amazon CloudWatch
- AWS X-Ray

REL11-BP05 Usa la stabilità statica per prevenire il comportamento bimodale

I carichi di lavoro devono essere staticamente stabili e funzionare in una singola modalità normale. Il comportamento bimodale si verifica quando il carico di lavoro presenta un comportamento diverso in modalità normale e in modalità di guasto.

Ad esempio, ciò potrebbe accadere nel momento in cui si prova a ripristinare un guasto nella zona di disponibilità avviando nuove istanze in una zona di disponibilità diversa. Questo approccio può comportare una risposta bimodale durante una modalità di guasto. È invece necessario creare carichi di lavoro che siano staticamente stabili e operino in una sola modalità. In questo esempio, le nuove istanze avrebbero dovuto essere allocate nella seconda zona di disponibilità già prima del guasto. Questo design staticamente stabile verifica che il carico di lavoro funzioni in una sola modalità.

Risultato desiderato: i carichi di lavoro non presentano un comportamento bimodale in modalità normale e in modalità di guasto.

Anti-pattern comuni:

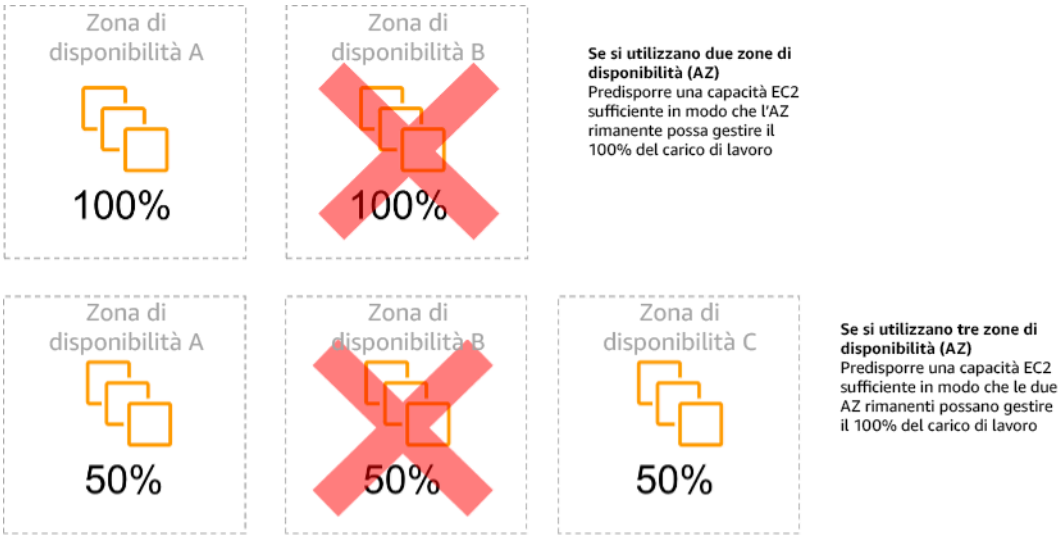
- Supporre che le risorse possano sempre essere allocate indipendentemente dall'ambito del guasto.
- Tentare di acquisire risorse in modo dinamico durante un guasto.
- Non rendere disponibili risorse adeguate tra zone o regioni diverse fino a quando non si verifica un guasto.
- Considerare i progetti staticamente stabili solo per risorse di calcolo.

Vantaggi dell'adozione di questa best practice: i carichi di lavoro eseguiti con progetti staticamente stabili sono in grado di avere risultati prevedibili durante eventi normali e di guasto.

Livello di rischio associato se questa best practice non fosse adottata: medio

Guida all'implementazione

Il comportamento bimodale ha luogo quando il carico di lavoro mostra un comportamento diverso in modalità normale e di guasto, ad esempio facendo affidamento sull'avvio di nuove istanze se una zona di disponibilità presenta un malfunzionamento. Un esempio di comportamento bimodale si verifica quando EC2 progetti Amazon stabili forniscono un numero sufficiente di istanze in ciascuna zona di disponibilità per gestire il carico di lavoro in caso di rimozione di una zona di disponibilità. Elastic Load Balancing o Amazon Route 53 controllano lo stato in modo da trasferire il carico dalle istanze danneggiate. Dopo lo spostamento del traffico, utilizzalo per sostituire in modo asincrono AWS Auto Scaling le istanze dalla zona in cui si è verificato l'errore e avviarle nelle zone funzionanti. La stabilità statica per l'implementazione dell'elaborazione (come EC2 istanze o contenitori) garantisce la massima affidabilità.



Stabilità statica delle EC2 istanze nelle zone di disponibilità

Questo approccio deve essere valutato rispetto al costo associato al modello e al valore aziendale attribuito al mantenimento della disponibilità del carico di lavoro in tutti i casi di resilienza. Fornire una minore capacità di elaborazione e affidarsi all'avvio di nuove istanze in caso di guasto è meno costoso. Tuttavia, in caso di guasti su larga scala, come una zona di disponibilità o un problema a livello regionale, tale approccio è meno efficace, perché si basa su un piano operativo e sulla disponibilità di risorse sufficienti nelle zone o nelle regioni non interessate dal problema.

La soluzione deve inoltre valutare l'affidabilità rispetto ai costi necessari per il carico di lavoro. Le architetture di stabilità statica si applicano a una varietà di architetture, tra cui istanze di calcolo distribuite tra zone di disponibilità, progetti di repliche di lettura del database, progetti di cluster Kubernetes (AmazonEKS) e architetture di failover multiregione.

È anche possibile implementare un progetto staticamente più stabile utilizzando più risorse in ciascuna zona. Aggiungendo più zone, si riduce la quantità di elaborazione aggiuntiva necessaria per la stabilità statica.

Un altro esempio di comportamento bimodale potrebbe derivare da un timeout di rete in grado di causare un tentativo di aggiornamento dello stato di configurazione dell'intero sistema. Ciò potrebbe aggiungere un carico imprevisto su un altro componente che potrebbe quindi generare un errore, innescando ulteriori conseguenze impreviste. Questo loop di feedback negativo influisce sulla disponibilità del carico di lavoro. Al contrario, è possibile creare sistemi che siano staticamente stabili e funzionino in una sola modalità. Un progetto staticamente stabile potrebbe eseguire con continuità un'attività e aggiornare sempre, con cadenza regolare, lo stato della configurazione. Quando una chiamata non va a buon fine, il carico di lavoro può utilizzare il valore precedentemente memorizzato nella cache e segnalare un allarme.

Un altro esempio di comportamento bimodale è consentire ai client di bypassare la cache del carico di lavoro quando si verificano guasti. Potrebbe sembrare una soluzione che soddisfa le esigenze del client, ma non dovrebbe essere consentita perché modifica in modo significativo le richieste sul carico di lavoro e potrebbe causare dei guasti.

Valuta i carichi di lavoro critici per determinare quali carichi di lavoro richiedono questo tipo di progettazione di resilienza. Per quelli considerati critici, deve essere esaminato ogni componente dell'applicazione. Alcuni tipi di servizi che richiedono valutazioni di stabilità statica sono:

- Elaborazione: AmazonEC2, EKS -EC2, ECS -EC2, EMR - EC2
- Database: Amazon Redshift, AmazonRDS, Amazon Aurora
- Archiviazione: Amazon S3 (zona singola), Amazon EFS (supporti), Amazon FSx (supporti)
- Bilanciatori del carico: in base a determinati progetti

Passaggi dell'implementazione

- Realizzare sistemi che siano staticamente stabili e operino in una sola modalità. In questo caso, effettuare il provisioning di un numero sufficiente di istanze in ogni zona o regione di disponibilità per gestire la capacità del carico di lavoro qualora venga rimossa una zona o regione

di disponibilità. Per l'indirizzamento verso risorse integre è possibile utilizzare una varietà di servizi, come:

- Routing tra regioni DNS
- MRAPRouting Amazon S3 MultiRegion
- AWS Global Accelerator
- Controller di ripristino delle applicazioni Amazon

- Configura repliche di lettura del database per tenere conto della perdita di una singola istanza primaria o di una replica di lettura. Se il traffico viene servito da repliche di lettura, la quantità in ogni zona di disponibilità e in ogni regione deve corrispondere al fabbisogno complessivo in caso di guasto della zona o della regione.
- Configurare i dati critici nel sistema di archiviazione Amazon S3 progettato per essere staticamente stabile rispetto ai dati archiviati in caso di guasto della zona di disponibilità. In caso di utilizzo della classe di archiviazione Amazon S3 One Zone-IA, questa non deve essere considerata staticamente stabile, poiché la perdita di tale zona riduce al minimo l'accesso ai dati archiviati.
- I bilanciatori del carico sono a volte configurati in modo errato o sono progettati per servire una zona di disponibilità specifica. In questo caso, il design staticamente stabile potrebbe consistere nel distribuire un carico di lavoro su più livelli AZs in un progetto più complesso. Il progetto originale potrebbe essere utilizzato per ridurre il traffico tra zone per motivi di sicurezza, latenza o costi.

Risorse

Best practice Well-Architected correlate:

- Definizione di disponibilità
- REL11-BP01 Monitora tutti i componenti del carico di lavoro per rilevare i guasti
- REL11-BP04 Fai affidamento sul piano dati e non sul piano di controllo durante il ripristino

Documenti correlati:

- Minimizing Dependencies in a Disaster Recovery Plan
- The Amazon Builders' Library: stabilità statica con le zone di disponibilità
- Limiti di isolamento dei guasti
- Stabilità statica con le zone di disponibilità
- Multi-zona RDS

- Minimizing Dependencies in a Disaster Recovery Plan
- Routing interregionale DNS
- MRAPRouting Amazon S3 MultiRegion
- AWS Global Accelerator
- Controller di ripristino delle applicazioni Amazon
- Amazon S3 a zona singola
- Bilanciamento del carico su più zone

Video correlati:

- Stabilità statica in AWS: AWS re:Invent 2019: Presentazione di The Amazon Builders' Library () DOP328

REL11-BP06 Invia notifiche quando gli eventi influiscono sulla disponibilità

Le notifiche vengono inviate al rilevamento del superamento delle soglie, anche se l'evento causato dal problema è stato risolto automaticamente.

Il ripristino automatizzato consente al carico di lavoro di risultare affidabile. Tuttavia, potrebbe anche nascondere problemi sottostanti che devono essere risolti. Implementa il monitoraggio e gli eventi appropriati in modo da poter rilevare i modelli di problemi, inclusi quelli risolti dalla diagnostica automatica e risolvere così i problemi della causa principale.

I sistemi resilienti sono progettati in modo che gli eventi di degrado vengano immediatamente comunicati ai team appropriati. Queste notifiche devono essere inviate tramite uno o più canali di comunicazione.

Risultato desiderato: gli avvisi vengono inviati immediatamente ai team operativi quando vengono superate le soglie, ad esempio i tassi di errore, la latenza o altri indicatori chiave di prestazione (KPI), in modo che questi problemi vengano risolti il prima possibile e l'impatto sugli utenti sia evitato o ridotto al minimo.

Anti-pattern comuni:

- Invio di un numero eccessivo di allarmi.
- Invio di allarmi non utilizzabili.

- Impostazione di soglie di allarme troppo alte (troppo sensibili) o troppo basse (troppo poco sensibili).
- Mancato invio di allarmi per dipendenze esterne.
- Mancata presa in considerazione dei guasti nell'area grigia nella progettazione di sistemi di monitoraggio e allarmi.
- Eseguire l'automazione del risanamento, ma senza avvisare il team competente che era necessario un intervento di ripristino.

Vantaggi derivanti dall'adozione di questa best practice: le notifiche di ripristino informano i team operativi e aziendali dei peggioramenti del servizio in modo che possano reagire immediatamente per ridurre al minimo sia il tempo medio di rilevamento (MTTD) che il tempo medio di riparazione (MTTR). Le notifiche degli eventi di ripristino consentono anche di non ignorare i problemi che si verificano di rado.

Livello di rischio associato se questa best practice non fosse adottata: medio. La mancata implementazione di meccanismi di monitoraggio e notifica degli eventi appropriati può comportare l'impossibilità di rilevare i modelli di problemi, compresi quelli risolti mediante la correzione automatica. Un team verrà informato del degrado del sistema solo nel momento in cui gli utenti contattano il servizio clienti o per caso.

Guida all'implementazione

Quando si definisce una strategia di monitoraggio, un allarme attivato è un evento comune. Questo evento dovrebbe contenere un identificatore dell'allarme, lo stato dell'allarme (ad esempio `IN ALARM` o `OK`) e i dettagli di ciò che lo ha attivato. In molti casi, è necessario rilevare un evento di allarme e inviare una notifica tramite e-mail. Questo è un esempio di operazione su un allarme. La notifica degli allarmi è fondamentale per l'osservabilità, in quanto informa le persone giuste della presenza di un problema. Tuttavia, quando le operazioni eseguite sulla base degli eventi raggiungono un certo grado di maturità nella soluzione di osservabilità, è possibile risolvere automaticamente il problema senza la necessità dell'intervento umano.

Una volta stabiliti gli allarmi di KPI monitoraggio, è necessario inviare avvisi ai team appropriati quando vengono superate le soglie. Tali avvisi possono essere utilizzati anche per attivare processi automatizzati che tenteranno di porre rimedio al danno o alla compromissione.

Per un monitoraggio delle soglie più complesso, è necessario prendere in considerazione gli allarmi compositi. Gli allarmi compositi utilizzano una serie di allarmi di KPI monitoraggio per creare un avviso basato sulla logica aziendale operativa. CloudWatchGli allarmi possono essere configurati

per inviare e-mail o per registrare gli incidenti in sistemi di tracciamento degli incidenti di terze parti utilizzando SNS l'integrazione di Amazon o Amazon. EventBridge

Passaggi dell'implementazione

Crea vari tipi di allarmi in base al modo in cui vengono monitorati i carichi di lavoro, ad esempio:

- Gli allarmi applicativi vengono utilizzati per rilevare quando una parte del carico di lavoro non funziona correttamente.
- Gli allarmi infrastrutturali indicano quando scalare le risorse. Gli allarmi possono essere visualizzati visivamente sulle dashboard, inviare avvisi tramite Amazon SNS o e-mail e utilizzare Auto Scaling per aumentare o ridurre le risorse del carico di lavoro.
- È possibile creare semplici allarmi statistici per monitorare quando una metrica supera una soglia statica per un numero specificato di periodi di valutazione.
- Gli allarmi compositi possono tenere conto di allarmi complessi provenienti da più fonti.
- Una volta creato l'allarme è possibile generare eventi di notifica appropriati. Puoi richiamare direttamente un Amazon SNS API per inviare notifiche e collegare qualsiasi automazione per la riparazione o la comunicazione.
- Integra il monitoraggio di Amazon Health Aware per consentire il monitoraggio della visibilità AWS delle risorse che potrebbero presentare un deterioramento. Per i carichi di lavoro aziendali essenziali, questa soluzione fornisce l'accesso ad avvisi proattivi e in tempo reale per i servizi. AWS

Risorse

Best practice Well-Architected correlate:

- Definizione di disponibilità

Documenti correlati:

- Creazione di un CloudWatch allarme basato su una soglia statica
- Che cos'è Amazon EventBridge?
- What is Amazon Simple Notification Service?
- Publishing Custom Metrics
- Utilizzo di Amazon CloudWatch Alarms

- Amazon Health Aware (AHA)
- Imposta CloudWatch allarmi compositi
- Cosa c'è di nuovo in AWS Observability a re:Invent 2022

Strumenti correlati:

- CloudWatch
- CloudWatch X-Ray

REL11-BP07 Progetta il tuo prodotto per soddisfare gli obiettivi di disponibilità e gli accordi sui livelli di servizio di uptime () SLAs

Progetta il tuo prodotto in modo da soddisfare gli obiettivi di disponibilità e gli accordi sui livelli di servizio di uptime (). SLAs Se pubblicate o accettate privatamente gli obiettivi di disponibilità o l'uptimeSLAs, verificate che l'architettura e i processi operativi siano progettati per supportarli.

Risultato desiderato: ogni applicazione ha un obiettivo definito per la disponibilità e SLA le metriche prestazionali, che possono essere monitorate e gestite per soddisfare i risultati aziendali.

Anti-pattern comuni:

- Progettazione e implementazione di carichi di lavoro senza impostarne alcuno. SLAs
- SLAle metriche sono troppo elevate senza motivazioni o requisiti aziendali.
- Impostazione SLAs senza tenere conto delle dipendenze e delle relative dipendenze sottostanti. SLA
- Progettazione delle applicazioni senza tenere conto del Modello di responsabilità condivisa per la resilienza.

Vantaggi dell'adozione di questa best practice: soddisfare gli obiettivi aziendali e le aspettative dei clienti grazie alla progettazione di applicazioni in base a obiettivi chiave in termini di resilienza. Questi obiettivi orientano un processo di progettazione delle applicazioni in grado di valutare diverse tecnologie e tenere conto di vari compromessi.

Livello di rischio associato se questa best practice non fosse adottata: medio

Guida all'implementazione

La progettazione delle applicazioni deve tenere conto di una serie eterogenea di requisiti derivati da obiettivi aziendali, operativi e finanziari. Nell'ambito dei requisiti operativi, i carichi di lavoro devono avere obiettivi specifici in termini di metriche di resilienza, in modo da poter essere monitorati e supportati correttamente. Le metriche di resilienza non devono essere impostate o derivate dopo l'implementazione del carico di lavoro. Devono invece essere definite durante la fase di progettazione e contribuire a determinare i diversi compromessi e decisioni.

- Ogni carico di lavoro deve avere una serie di metriche di resilienza propria. Le metriche possono essere diverse da quelle di altre applicazioni aziendali.
- La riduzione delle dipendenze può avere un impatto positivo sulla disponibilità. Ogni carico di lavoro deve considerare le sue dipendenze e le relative. SLAs In generale, seleziona dipendenze con obiettivi di disponibilità uguali o maggiori rispetto agli obiettivi del carico di lavoro.
- Prendi in considerazione progettazioni con accoppiamento debole in modo che il carico di lavoro possa funzionare correttamente anche in caso di dipendenze compromesse, se possibile.
- Riduci le dipendenze del piano di controllo (control-plane), in particolare durante un ripristino o un peggioramento delle prestazioni. Valuta le progettazioni staticamente stabili per carichi di lavoro mission critical. Usa il contenimento delle risorse per aumentare la disponibilità delle dipendenze in un carico di lavoro.
- L'osservabilità e la strumentazione sono fondamentali per ridurre il tempo medio SLAs di rilevamento (MTTD) e il tempo medio di riparazione (). MTTR
- Guasti meno frequenti (più lunghiMTBF), tempi di rilevamento dei guasti più brevi (più breviMTTD) e tempi di riparazione più brevi (più breviMTTR) sono i tre fattori utilizzati per migliorare la disponibilità nei sistemi distribuiti.
- La definizione e l'applicazione di metriche di resilienza per un carico di lavoro sono essenziali per qualsiasi progettazione efficace. Queste progettazioni devono tenere conto dei compromessi introdotti dalla complessità di progettazione, delle dipendenze dei servizi, delle prestazioni, del dimensionamento e dei costi.

Passaggi dell'implementazione

- Esamina e documenta la progettazione del carico di lavoro cercando di rispondere alle domande seguenti:
  - Dove vengono usati i piani di controllo (control-plane) nel carico di lavoro?
  - Come viene implementata la tolleranza ai guasti nel carico di lavoro?

  - Quali sono i modelli di progettazione per dimensionamento, scalabilità automatica, ridondanza e componenti a disponibilità elevata?
  - Quali sono i requisiti per la disponibilità e la coerenza dei dati?
  - Vi sono aspetti da considerare in fatto di contenimento delle risorse o stabilità statica delle risorse?
  - Quali sono le dipendenze dei servizi?
- Definisci le SLA metriche in base all'architettura del carico di lavoro mentre lavori con le parti interessate. Considera tutte le SLAs dipendenze utilizzate dal carico di lavoro.
- Una volta impostato l'SLAobiettivo, ottimizza l'architettura per soddisfare i. SLA
- Una volta definito il progettoSLA, ciò consentirà di implementare le modifiche operative, l'automazione dei processi e i runbook che si concentreranno anche sulla riduzione MTTD eMTTR.
- Una volta implementato, monitora e segnala su. SLA

Risorse

Best practice correlate:

- REL03-BP01 Scegli come segmentare il tuo carico di lavoro
- REL10-BP01 Implementazione del carico di lavoro in più sedi
- REL11-BP01 Monitora tutti i componenti del carico di lavoro per rilevare i guasti
- REL11-BP03 Automatizza la guarigione su tutti i livelli
- REL12-BP05 Verifica la resilienza utilizzando l'ingegneria del caos
- REL13-BP01 Definire gli obiettivi di ripristino per tempi di inattività e perdita di dati
- Comprendere lo stato del carico di lavoro

Documenti correlati:

- Availability with redundancy
- Pilastro dell'affidabilità: disponibilità
- Measuring availability
- AWS Limiti di isolamento dei guasti
- Modello di responsabilità condivisa per la resilienza
- Stabilità statica con le zone di disponibilità

- AWS Accordi sui livelli di servizio (SLAs)
- Linee guida per l'architettura basata su celle su AWS
- AWS infrastruttura
- Whitepaper Advanced Multi-AZ Resiliance Patterns

Servizi correlati:

- Amazon CloudWatch
- AWS Config
- AWS Trusted Advisor

REL12. Come si testa l'affidabilità?

Dopo aver progettato il carico di lavoro in modo da essere resiliente alle sollecitazioni della produzione, i test sono l'unico modo per verificare il funzionamento corretto e offrire la resilienza prevista.

Best practice

- REL12-BP01 Usa i playbook per indagare sui guasti
- REL12-BP02 Eseguire l'analisi post-incidente
- REL12-Requisiti funzionali del test -BP03
- REL12-BP04 Test di scalabilità e requisiti prestazionali
- REL12-BP05 Verifica la resilienza utilizzando l'ingegneria del caos
- REL12-BP06 Conduci regolarmente giornate di gioco

REL12-BP01 Usa i playbook per indagare sui guasti

Consenti risposte coerenti e tempestive a scenari di guasto che non sono ben compresi, documentando il processo di analisi nei playbook. I playbook sono le fasi predefinite eseguite per identificare i fattori che contribuiscono a uno scenario di guasto. I risultati provenienti da un passaggio del processo vengono utilizzati per stabilire i passaggi successivi da intraprendere fino all'identificazione o alla risoluzione del problema.

Il playbook è una pianificazione proattiva che è necessario eseguire, in modo da potere intraprendere azioni reattive in modo efficace. Se durante la produzione si verificano scenari di guasto non coperti

dal playbook, risolvi innanzitutto il problema (spegni l'incendio). Quindi torna indietro e osserva le fasi intraprese per risolvere il problema e utilizzale per aggiungere una nuova voce al playbook.

Tieni presente che i playbook vengono utilizzati in risposta a specifici incidenti, mentre i runbook vengono utilizzati per ottenere esiti specifici. Spesso, i runbook vengono utilizzati per le attività di routine e i playbook vengono utilizzati per rispondere a eventi non di routine.

Anti-pattern comuni:

- Pianificare la distribuzione di un carico di lavoro senza conoscere i processi per diagnosticare i problemi o rispondere agli incidenti.
- Decisioni non pianificate sui sistemi da cui raccogliere log e parametri durante l'analisi di un evento.
- Non conservare parametri e eventi abbastanza a lungo da poter recuperare i dati.

Vantaggi dell'adozione di questa best practice: l'acquisizione dei playbook garantisce l'esecuzione coerente dei processi. La codifica dei playbook limita l'introduzione di errori derivanti dall'attività manuale. L'automazione dei playbook riduce il tempo necessario per rispondere a un evento eliminando il requisito per l'intervento dei membri del team o fornendo loro informazioni aggiuntive quando inizia l'intervento.

Livello di rischio associato se questa best practice non fosse adottata: elevato

Guida all'implementazione

- Utilizza playbook per identificare i problemi. I playbook sono processi documentati per eseguire indagini sui problemi. Promuovi risposte coerenti e tempestive agli scenari di errore documentando i processi nei playbook. I playbook devono contenere le informazioni e le istruzioni necessarie affinché una persona adeguatamente qualificata possa raccogliere le informazioni applicabili, identificare potenziali fonti di errore, isolare i guasti e stabilire i fattori che contribuiscono all'origine di un problema (eseguire l'analisi post-incidente).
  - Implementazione dei playbook come codice. Esegui le operazioni come codice mediante lo scripting dei playbook per assicurare coerenza e ridurre gli errori causati dai processi manuali. I playbook possono essere composti da più script che rappresentano le diverse fasi che potrebbero essere necessarie per identificare i fattori che contribuiscono all'origine di un problema. Le attività dei runbook possono essere richiamate o eseguite nell'ambito delle attività dei playbook oppure possono richiedere l'esecuzione di un playbook in risposta agli eventi identificati.
    - Automatizza i tuoi playbook operativi con Systems Manager AWS

- AWS Comando di esecuzione di Systems Manager
- AWS Systems Manager Automation
- Che cos'è AWS Lambda?
- Che cos'è Amazon EventBridge?
- Utilizzo di Amazon CloudWatch Alarms

Risorse

Documenti correlati:

- AWS Systems Manager Automation
- AWS Comando di esecuzione di Systems Manager
- Automatizza i tuoi playbook operativi con Systems Manager AWS
- Utilizzo di Amazon CloudWatch Alarms
- Utilizzo di Canaries (Amazon CloudWatch Synthetics)
- Che cos'è Amazon EventBridge?
- Che cos'è AWS Lambda?

Esempi correlati:

- Automazione delle operazioni con playbook e runbook

REL12-BP02 Eseguire l'analisi post-incidente

Esamina gli eventi che influiscono sui clienti e identifica i fattori che vi hanno contribuito e gli elementi di azione preventivi. Utilizza queste informazioni per sviluppare modi per limitare o prevenire il ripetersi degli incidenti. Sviluppa procedure per attivare risposte rapide ed efficaci. Comunica i fattori che hanno contribuito al presentarsi dell'imprevisto e le azioni correttive secondo necessità, specificamente mirate per il pubblico di destinazione. All'occorrenza, adotta un metodo per comunicare queste cause ad altri.

Valuta perché i test esistenti non hanno individuato il problema. Aggiungi i test per questo caso se i test non esistono già.

Risultato desiderato: i tuoi team dispongono di un approccio coerente e concordato per la gestione dell'analisi post-incidente. Un meccanismo è il processo di correzione dell'errore (). COE Il COE

processo aiuta i team a identificare, comprendere e affrontare le cause profonde degli incidenti, oltre a creare meccanismi e barriere per limitare la probabilità che lo stesso incidente si ripeta.

Anti-pattern comuni:

- Individuare i fattori che hanno contribuito al verificarsi dell'incidente, ma non continuare a cercare in maniera più approfondita altri potenziali problemi e approcci da mitigare.
- Identificare le cause degli errori umani senza fornire alcuna formazione o automazione che potrebbe prevenirli.
- Concentrarsi sull'attribuzione delle colpe piuttosto che sulla comprensione della causa principale, creando così una cultura della paura e ostacolando la comunicazione costruttiva
- Mancata condivisione delle informazioni, che mantiene gli esiti dell'analisi degli incidenti all'interno di un gruppo ristretto e impedisce ad altri di beneficiare delle lezioni apprese
- Nessun meccanismo che consenta di acquisire le conoscenze formali; in questo modo si perdono informazioni preziose in quanto non vengono preservate le lezioni apprese sotto forma di best practice aggiornate, con il conseguente rischio che gli incidenti si ripetano con la stessa causa principale o causa simile

Vantaggi dell'adozione di questa best practice: l'esecuzione di analisi post-incidente e la condivisione dei risultati consente ad altri carichi di lavoro di mitigare il rischio se hanno implementato gli stessi fattori che hanno contribuito al verificarsi dell'incidente e permette loro di implementare la mitigazione o il ripristino automatico prima che si verifichi un incidente.

Livello di rischio associato se questa best practice non fosse adottata: elevato

Guida all'implementazione

Una buona analisi post-incidente fornisce opportunità per proporre soluzioni comuni a problemi con modelli di architettura utilizzati in altri punti nei tuoi sistemi.

Una pietra miliare del COE processo è la documentazione e la risoluzione dei problemi. È consigliabile definire un modo standard per documentare le cause principali critiche e assicurarsi che queste vengano esaminate e risolte. Assegna in modo chiaro il responsabile del processo di analisi post-incidente. Nomina un team o una persona responsabile della supervisione delle indagini e dei follow-up degli incidenti.

Promuovi una cultura basata sull'apprendimento e sul miglioramento piuttosto che sull'attribuzione di colpe. Insisti sul fatto che l'obiettivo è prevenire incidenti futuri e non penalizzare le persone.

Sviluppa procedure ben definite per l'esecuzione delle analisi post-incidente. Queste procedure dovrebbero stabilire le misure da adottare, le informazioni da raccogliere e le questioni chiave da risolvere durante l'analisi. Svolgi indagini approfondite sugli incidenti, andando oltre le cause immediate per identificare le cause principali e i fattori determinanti. Utilizza tecniche come i Cinque Perché per analizzare in modo approfondito i problemi sottostanti.

Mantieni un archivio delle conclusioni derivanti dalle analisi degli incidenti. Queste conoscenze formali possono fungere da riferimento per futuri incidenti e attività di prevenzione. Condividi gli esiti e gli approfondimenti delle analisi post-incidente e valuta la possibilità di organizzare riunioni di revisione post-incidente con invito aperto per discutere i risultati e le conclusioni.

Passaggi dell'implementazione

- Durante l'analisi post-incidente, assicurati che il processo non comporti la colpevolizzazione delle parti coinvolte. Ciò consente alle parti interessate di essere imparziali rispetto delle azioni correttive proposte, nonché di promuovere l'autovalutazione e la collaborazione a livello di team.
- Definisci una procedura standardizzata per documentare i problemi critici. Una struttura di esempio per tale documento è la seguente:
  - Che cos'è successo?
  - Quale impatto ha avuto su clienti e attività?
  - Qual è stata la causa principale?
  - Di quali dati disponi a supporto di ciò?
    - Ad esempio, metriche e grafici
  - Quali sono state le principali implicazioni sui pilastri critici, specialmente per quanto riguarda la sicurezza?
    - Quando progetti l'architettura dei carichi di lavoro, devi trovare dei compromessi tra i pilastri su cui si regge il contesto aziendale. Le decisioni aziendali possono stabilire le priorità di progettazione. Potresti ridurre i costi a spese dell'affidabilità in ambienti di sviluppo oppure, per quanto riguarda le soluzioni mission-critical, potresti ottimizzare l'affidabilità con costi maggiori. La sicurezza ha la massima priorità quando si tratta di proteggere i tuoi clienti.
  - Quali lezioni hai imparato?
  - Quali azioni correttive stai adottando?
    - Elementi d'azione
    - Voci correlate
- Crea precise procedure operative standard per lo svolgimento delle analisi post-incidente.

- Configura un processo standardizzato di segnalazione degli incidenti. Documenta in modo esaustivo tutti gli incidenti, includendo il rapporto iniziale sull'incidente, i log, le comunicazioni e le azioni intraprese durante l'incidente.
- Ricorda che un incidente non necessariamente comporta un'interruzione del servizio. Potrebbe trattarsi di un near miss o di un sistema che funziona in modo imprevisto pur continuando a svolgere la sua funzione aziendale.
- Migliora continuamente il processo di analisi post-incidente sulla base dei feedback e delle lezioni apprese.
- Acquisisci gli esiti chiave in un sistema di gestione delle conoscenze e valuta eventuali modelli da aggiungere alle linee guide per gli sviluppatori o alle liste di controllo usate nella fase di pre-implementazione.

Risorse

Documenti correlati:

- Perché dovresti sviluppare una correzione dell'errore () COE

Video correlati:

- Amazon's approach to failing successfully
- AWS re:Invent 2021 - Amazon Builders' Library: eccellenza operativa in Amazon

REL12-Requisiti funzionali del test -BP03

Utilizza tecniche come i test di unità e i test di integrazione per convalidare le funzionalità richieste.

Puoi ottenere i migliori risultati quando questi test vengono eseguiti automaticamente come parte delle operazioni di sviluppo e implementazione. Ad esempio, utilizzando AWS CodePipeline, gli sviluppatori inseriscono le modifiche in un repository di origine in cui CodePipeline rileva automaticamente le modifiche. Queste modifiche vengono create e vengono eseguiti test. Dopo aver completato i test, il codice compilato viene distribuito ai server della gestione temporanea per il test. Dal server di staging, CodePipeline esegue più test, come test di integrazione o di carico. Una volta completati con successo tali test, CodePipeline distribuisce il codice testato e approvato nelle istanze di produzione.

Inoltre, l'esperienza dimostra che il test sintetico delle transazioni (noto anche come "test canary", da non confondere con le distribuzioni canary) in grado di eseguire e simulare il comportamento dei clienti è uno dei processi di test più importanti. Esegui questi test costantemente sugli endpoint del carico di lavoro da diverse posizioni remote. Amazon CloudWatch Synthetics ti consente di creare canaries per monitorare i tuoi endpoint e. APIs

Livello di rischio associato se questa best practice non fosse adottata: elevato

Guida all'implementazione

- Test dei requisiti funzionali. Includono test di unità e test di integrazione che convalidano la funzionalità richiesta.
  - Usalo CodePipeline con AWS CodeBuild per testare il codice ed eseguire build
  - AWS CodePipeline Aggiunge il supporto per i test di integrazione unitari e personalizzati con AWS CodeBuild
  - Continuous Delivery and Continuous Integration
  - Utilizzo di Canaries (Amazon CloudWatch Synthetics)
  - Automazione e test del software

Risorse

Documenti correlati:

- APNPartner: partner che possono contribuire all'implementazione di una pipeline di integrazione continua
- AWS CodePipeline Aggiunge il supporto per i test di integrazione unitari e personalizzati con AWS CodeBuild
- Marketplace AWS: prodotti utilizzabili per l'integrazione continua
- Continuous Delivery and Continuous Integration
- Automazione e test del software
- Usalo CodePipeline con AWS CodeBuild per testare il codice ed eseguire build
- Utilizzo di Canaries (Amazon CloudWatch Synthetics)

REL12-BP04 Test di scalabilità e requisiti prestazionali

Utilizza tecniche come i test di carico per convalidare che il carico di lavoro soddisfi i requisiti di dimensionamento e prestazioni.

Nel cloud, puoi creare un ambiente di test su scala di produzione on demand per il tuo carico di lavoro. Se esegui questi test su un'infrastruttura ridotta verticalmente, devi scalare i risultati osservati in base a ciò che pensi accadrà in produzione. I test di carico e prestazioni possono essere eseguiti anche in produzione se si fa attenzione a non influire sugli utenti effettivi e si contrassegna con tag i dati di test in modo da non utilizzare dati utente reali e non danneggiare le statistiche di utilizzo o i report di produzione.

Con i test, ti assicuri che le risorse di base, le impostazioni di dimensionamento, le quote di servizio e la progettazione di resilienza funzionino come previsto sotto carico.

Livello di rischio associato se questa best practice non fosse adottata: elevato

Guida all'implementazione

- Test dei requisiti di dimensionamento e prestazioni. Esegui test del carico per verificare che il carico di lavoro soddisfi i requisiti di dimensionamento e prestazioni.
  - Test di carico distribuito su AWS: simula migliaia di utenti connessi
  - Apache JMeter
    - Implementa la tua applicazione in un ambiente identico al tuo ambiente di produzione ed esegui un test di carico.
      - Utilizza un'infrastruttura come code concept per creare un ambiente il più simile possibile al tuo ambiente di produzione.

Risorse

Documenti correlati:

- Test di carico distribuito su AWS: simula migliaia di utenti connessi
- Apache JMeter

REL12-BP05 Verifica la resilienza utilizzando l'ingegneria del caos

Esegui regolarmente esperimenti di ingegneria del caos in ambienti di produzione o per quanto possibile ambienti analoghi per capire in che modo il sistema risponde a condizioni avverse.

Risultato desiderato:

La resilienza del carico di lavoro viene regolarmente verificata mediante l'applicazione dell'ingegneria del caos sotto forma di esperimenti di iniezione di guasti o di inserimento di carichi imprevisti, nonché mediante il test della resilienza che convalida i comportamenti previsti noti del carico di lavoro durante un evento. Combina l'ingegneria del caos e i test della resilienza per verificare se il carico di lavoro è in grado di superare i guasti dei componenti ed eseguire il ripristino da interruzioni del servizio impreviste con un impatto minimo o nullo.

Anti-pattern comuni:

- Progettazione della resilienza, ma mancata verifica del funzionamento del carico di lavoro nel suo complesso in caso di errori.
- Mancata sperimentazione in scenari reali e con carichi previsti.
- Mancato trattamento degli esperimenti come codice o loro conservazione durante il ciclo di sviluppo.
- Mancata esecuzione degli esperimenti di ingegneria del caos sia nella pipeline CI/CD che esternamente alle implementazioni.
- Mancato utilizzo delle precedenti analisi post-incidente durante la determinazione degli errori su cui eseguire i test.

Vantaggi dell'adozione di questa best practice: l'introduzione di errori per verificare la resilienza del carico di lavoro consente di verificare che le procedure di ripristino della progettazione resiliente funzionerà se viene generato un vero e proprio errore.

Livello di rischio associato se questa best practice non fosse adottata: medio

Guida all'implementazione

L'ingegneria del caos offre ai team la possibilità di continuare a inserire scenari di errore reali (simulazioni) in modo controllato a livello di fornitore di servizi, infrastruttura, carico di lavoro e componente con un impatto minimo o nullo per i clienti. Consente inoltre ai team di imparare dagli errori e osservare, misurare e migliorare la resilienza dei carichi di lavoro, nonché verificare l'attivazione degli avvisi e se tali avvisi vengono recapitati ai team se si verifica un evento definito.

Se applicata in modo continuativo, l'ingegneria del caos può mettere in evidenza i difetti del carico di lavoro che, se non risolti, possono avere ripercussioni negative sulla disponibilità e sulle operazioni.

> **Note**
>
> L'ingegneria del caos è la disciplina che sperimenta un sistema per creare fiducia nella capacità del sistema di affrontare condizioni turbolenti nella produzione. – Principles of Chaos Engineering

Se un sistema è in grado di sopportare queste interruzioni, l'esperimento di ingegneria del caos deve essere convertito in test automatico di regressione. In questo modo, gli esperimenti sul caos devono essere eseguiti come parte del ciclo di vita di sviluppo dei sistemi (SDLC) e come parte della pipeline CI/CD.

Per garantire che il carico di lavoro sia in grado di gestire un guasto del componente, esegui l'iniezione di eventi di errore reali durante l'esecuzione degli esperimenti. Ad esempio, sperimenta la perdita di EC2 istanze Amazon o il failover dell'istanza di RDS database Amazon principale e verifica che il carico di lavoro non ne risenta (o lo sia solo in minima parte). Utilizza una combinazione di errori dei componenti per simulare gli eventi che possono essere causati da un'interruzione del servizio in una zona di disponibilità.

Per i guasti a livello di applicazione (come i crash), puoi iniziare con fattori di stress come la memoria e l'esaurimento. CPU

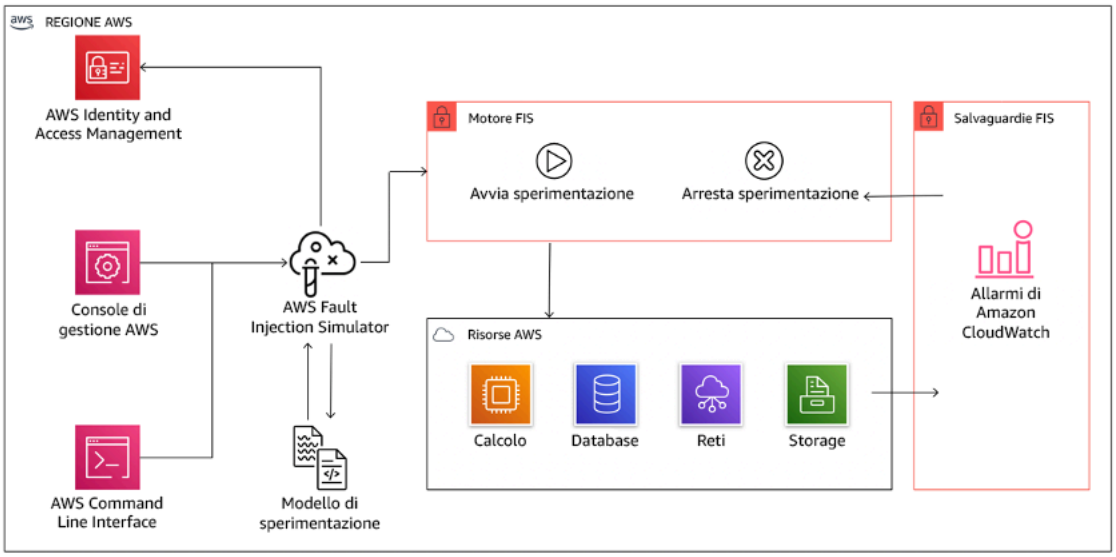
Per convalidare i meccanismi di fallback o failover per le dipendenze esterne causate da interruzioni intermittenti dei servizi di rete, i componenti devono simulare tale evento bloccando l'accesso ai fornitori di terze parti per una durata specificata, che può durare da pochi secondi ad alcune ore.

Altre modalità di degrado possono causare funzionalità ridotte e risposte lente, spesso con conseguente interruzione dei servizi. Le fonti comuni di questo degrado sono una maggiore latenza nei servizi critici e una comunicazione di rete inaffidabile (pacchetti persi). Gli esperimenti con questi errori, inclusi effetti di rete come latenza, messaggi persi ed DNS errori, potrebbero includere l'impossibilità di risolvere un nome, accedere al servizio o stabilire connessioni a servizi dipendenti. DNS

Strumenti dell'ingegneria del caos:

AWS Fault Injection Service (AWS FIS) è un servizio completamente gestito per l'esecuzione di esperimenti di iniezione dei guasti che può essere utilizzato come parte della pipeline CD o al di fuori della pipeline. AWS FIS è un'ottima scelta da usare durante le giornate dedicate ai giochi di Chaos Engineering. Supporta l'introduzione simultanea di errori su diversi tipi di risorse tra cui

AmazonEC2, Amazon Elastic Container Service (AmazonECS), Amazon Elastic Kubernetes Service (Amazon) e EKS Amazon. RDS Questi errori includono l'interruzione delle risorse, il failover forzato, lo stress della memoria, la limitazione, la latenza e la perdita di CPU pacchetti. Poiché è integrato con Amazon CloudWatch Alarms, puoi impostare condizioni di arresto come guardrail per annullare un esperimento se causa un impatto imprevisto.



AWS Fault Injection Service si integra con AWS le risorse per consentirti di eseguire esperimenti di iniezione dei guasti per i tuoi carichi di lavoro.

Esistono anche diverse opzioni di terze parti per gli esperimenti di iniezione di guasti. Queste opzioni comprendono strumenti open source come Chaos Toolkit, Chaos Mesh e Litmus Chaos, nonché altre opzioni commerciali come Gremlin. Per ampliare la gamma di errori su cui è possibile iniettare, si AWS FIS integra con Chaos Mesh e Litmus Chaos AWS, che consentono di coordinare i flussi di lavoro di iniezione dei guasti tra più strumenti. Ad esempio, è possibile eseguire uno stress test su un pod CPU utilizzando gli errori Chaos Mesh o Litmus mentre si termina una percentuale selezionata casualmente di nodi del cluster utilizzando azioni di errore. AWS FIS

Passaggi dell'implementazione

1. Determinazione dei guasti da utilizzare per gli esperimenti.

   Valutazione della progettazione del carico di lavoro a livello di resilienza. Tali progettazioni (create seguendo le best practice del Framework Well-Architected) tengono conto dei rischi basati su dipendenze critiche, eventi passati, problemi noti e requisiti di conformità. Elenca i singoli elementi della progettazione che devono conservare la resilienza e gli errori per mitigare i quali è stata sviluppata. Per ulteriori informazioni sulla creazione di tali elenchi, consulta il whitepaper

sulla prontezza operativa, che illustra come creare un processo finalizzato alla prevenzione del ripetersi di incidenti precedenti. Il processo Failure Modes and Effects Analysis (FMEA) fornisce un framework per eseguire un'analisi a livello di componente dei guasti e del loro impatto sul carico di lavoro. FMEAè descritto più dettagliatamente da Adrian Cockcroft in Failure Modes and Continuous Resilience.

2. Assegna una priorità a ogni errore.

   Comincia con una categorizzazione approssimativa, ad esempio alta, media o bassa. Per valutare la priorità, considera la frequenza dell'errore e l'impatto dell'errore sul carico di lavoro nel suo complesso.

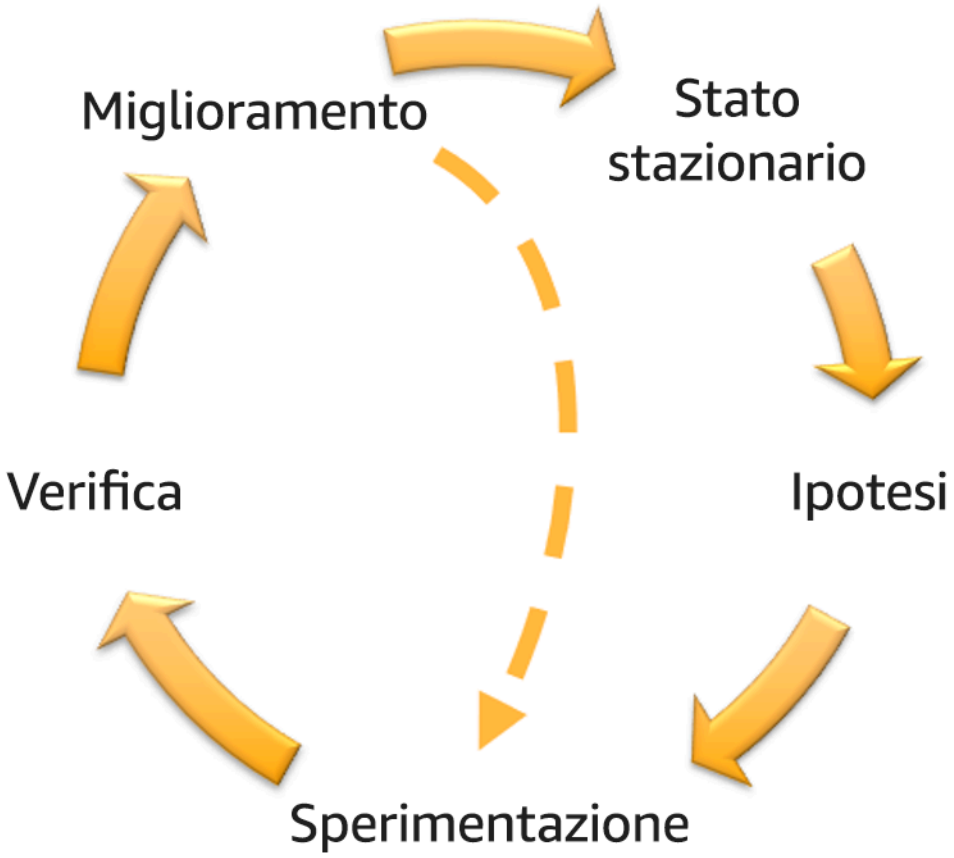
   Durante la valutazione della frequenza di un errore specifico, analizza i precedenti dati per lo stesso carico di lavoro, se disponibili. Se non sono disponibili, utilizza i dati di altri carichi di lavoro eseguiti in un ambiente simile.

   Durante la valutazione dell'impatto di un errore specifico, in genere maggiore è l'ambito dell'errore, maggiore sarà l'impatto. Considera la progettazione e lo scopo del carico di lavoro. Ad esempio, la capacità di accedere ai datastore di origine è di cruciale importanza per un carico di lavoro responsabile della trasformazione e dell'analisi dei dati. In questo caso, darai la precedenza agli esperimenti relativi agli errori di accesso, nonché a quelli con limitazione (della larghezza di banda della rete) e inserimento di latenza.

   Le analisi post-incidente rappresentano un'ottima fonte di dati per la comprensione della frequenza e dell'impatto delle modalità di errore.

   Utilizza la priorità assegnata per determinare il primo errore su cui eseguire l'esperimento e l'ordine in cui sviluppare i nuovi esperimenti di iniezione di guasti.

3. Per ogni esperimento eseguito, attieniti ai principi del volano dell'ingegneria del caos e della resilienza continua nella figura seguente.



Volano dell'ingegneria del caos e della resilienza continua, che utilizza il metodo scientifico di Adrian Hornsby.

a. Definisci lo stato stazionario come output misurabile di un carico di lavoro che indica un comportamento normale.

   Il carico di lavoro è associato allo stato stazionario se il suo funzionamento è affidabile e conforme a quanto previsto. Verifica pertanto che il carico di lavoro sia integro prima di definire lo stato stazionario. Lo stato stazionario non necessariamente indica l'assenza di impatto sul carico di lavoro se si verifica un errore in quanto una data percentuale di errori può rientrare nei limiti di valori accettabili. Lo stato stazionario rappresenta il punto di riferimento che verrà osservato durante l'esperimento e che metterà in evidenza le anomalie se le ipotesi definite nel passaggio successivo non sono conformi alle previsioni.

Ad esempio, lo stato stazionario di un sistema di pagamenti può essere definito come l'elaborazione di 300 TPS con una percentuale di successo del 99% e un tempo di andata e ritorno di 500 ms.

b. Definisci un'ipotesi in merito alle reazioni del carico di lavoro all'errore.

Un'ipotesi ottimale fa riferimento al modo in cui il carico di lavoro presumibilmente è in grado di ridurre l'impatto dell'errore e salvaguardare lo stato stazionario. Nell'ipotesi è definito che, dato un errore di un tipo specifico, il sistema o il carico di lavoro rimarrà nello stato stazionario poiché la progettazione del carico di lavoro ha previsto sistemi specifici di attenuazione degli errori. Il tipo di errore specifico e i sistemi di attenuazione devono essere specificati nell'ipotesi.

Per l'ipotesi è possibile utilizzare il seguente modello, anche se è accettabile una formulazione diversa:

> **Note**
>
> Se *`specific fault`* si verifica, il *`workload name`* il carico di lavoro sarà *`describe mitigating controls`* mantenere *`business or technical metric impact`*.

Per esempio:

- Se il 20% dei nodi del EKS gruppo di nodi Amazon viene disattivato, Transaction Create API continua a soddisfare il 99° percentile di richieste in meno di 100 ms (stato stazionario). EKSI nodi Amazon verranno ripristinati entro cinque minuti e i pod riceveranno il traffico pianificato ed elaborato entro otto minuti dall'inizio dell'esperimento. Gli avvisi verranno attivati entro tre minuti.
- Se si verifica un errore di una singola EC2 istanza Amazon, il controllo dello stato di Elastic Load Balancing del sistema di ordinazione farà sì che Elastic Load Balancing invii le richieste solo alle istanze integre rimanenti, mentre Amazon EC2 Auto Scaling sostituisce l'istanza fallita, mantenendo un aumento inferiore allo 0,01% degli errori lato server (5xx) (stato stazionario).
- Se l'istanza principale del RDS database Amazon si guasta, il carico di lavoro di raccolta dati Supply Chain eseguirà il failover e si connetterà all'istanza di database RDS Amazon in standby per mantenere meno di 1 minuto di errori di lettura o scrittura del database (stato stazionario).

c. Esegui l'esperimento inserendo l'errore.

Per impostazione predefinita, un esperimento deve essere a prova di errore e tollerato dal carico di lavoro. Se sei consapevole del fatto che il carico di lavoro avrà esito negativo, non eseguire l'esperimento. L'ingegneria del caos deve essere utilizzata per individuare scenari noti sconosciuti o scenari completamente sconosciuti. Per scenari noti sconosciuti si intendono gli scenari di cui sei consapevole ma che non comprendi appieno, mentre scenari completamente sconosciuti si riferiscono a quegli scenari a te non noti e che non comprendi appieno. L'esecuzione di esperimenti su un carico di lavoro non funzionante non può fornire nuovi approfondimenti chiarificatori. L'esperimento deve infatti essere pianificato con attenzione, essere caratterizzato da un ambito ben definito relativamente al suo impatto, nonché fornire un meccanismo di rollback applicabile in caso di esiti negativi imprevisti. Se il criterio di due diligence indica che il carico di lavoro è in grado di sostenere l'esperimento, procedi ed esegui l'esperimento. Sono disponibili varie opzioni per l'inserimento degli errori. Per i carichi di lavoro su AWS, AWS FIS offre diverse simulazioni di guasto predefinite denominate operazioni. Puoi anche definire azioni personalizzate da eseguire utilizzando documenti. AWS FISAWS Systems Manager

È sconsigliato l'uso di script personalizzati per gli esperimenti di ingegneria del caos, a meno che gli script non siano in grado di rilevare lo stato corrente del carico di lavoro, generare log e fornire meccanismi di rollback e condizioni di arresto, laddove possibile.

Un framework o set di strumenti efficace che supporta l'ingegneria del caos deve tenere traccia dello stato corrente di un esperimento, generare log e fornire meccanismi di rollback a supporto dell'esecuzione controllata di un esperimento. Inizia con un servizio consolidato come AWS FIS questo, che ti consenta di eseguire esperimenti con un ambito ben definito e meccanismi di sicurezza che annullino l'esperimento se l'esperimento introduce turbolenze impreviste. Per ulteriori informazioni su una più ampia varietà di esperimenti di utilizzo AWS FIS, consulta anche il laboratorio Resilient and Well-Architected Apps with Chaos Engineering. Inoltre, AWS Resilience Hub analizzerà il carico di lavoro e creerà gli esperimenti che potrai scegliere di implementare ed eseguire in AWS FIS.

> **Note**
>
> Per ogni esperimento, devi essere consapevole del suo ambito e del relativo impatto. È consigliabile eseguire la simulazione dell'errore in un ambiente non di produzione prima di eseguirla in un ambiente di produzione vero e proprio.

Gli esperimenti andrebbero eseguiti in produzione con un carico reale mediante distribuzioni canary che attivano l'implementazione di sistemi sperimentali e di controllo, laddove possibile. L'esecuzione degli esperimenti durante gli orari non di punta è altamente consigliata al fine di ridurre al massimo potenziali eventi negativi durante la prima esecuzione dell'esperimento negli ambienti di produzione. Inoltre, se l'utilizzo dell'effettivo traffico clienti costituisce un rischio eccessivo, puoi eseguire gli esperimenti utilizzando una sintesi del traffico nell'infrastruttura di produzione utilizzando implementazioni sperimentali e di controllo. Se l'utilizzo di un ambiente di produzione non è possibile, esegui gli esperimenti in ambienti di pre-produzione il più simili possibile agli effettivi ambienti di produzione.

Devi definire e monitorare i guardrail per essere sicuro che l'esperimento non abbia un impatto sul traffico di produzione o sugli altri sistemi che superi i limiti accettabili. Definisci condizioni di arresto per interrompere l'esperimento se viene raggiunta la soglia definita nella metrica del guardrail. In tali condizioni devono essere incluse le metriche relative allo stato stazionario del carico di lavoro e le metriche riferite ai componenti in cui inserisci l'errore. Un monitoraggio sintetico (definito anche canary utente) è una metrica che in genere deve essere inclusa come proxy utente. Le condizioni di arresto per AWS FIS sono supportate nel modello di esperimento, nella misura di un massimo di cinque condizioni di arresto per modello.

Uno dei principi dell'ingegneria del caos prevede la riduzione dell'ambito dell'esperimento e del relativo impatto.

Se da un lato deve essere prevista la possibilità di un determinato impatto negativo a breve termine, dall'altro il contenimento e la riduzione delle conseguenze negative degli esperimenti sono una responsabilità esclusiva dell'addetto all'ingegneria del caos.

Un metodo per verificare l'ambito e il potenziale impatto prevede l'esecuzione dell'esperimento dapprima in un ambiente non di produzione, la verifica che le soglie delle condizioni di arresto vengano attivate come previsto durante lo svolgimento di un esperimento e l'utilizzo effettivo delle misure di osservabilità finalizzate all'acquisizione di un'eccezione, anziché eseguire l'esperimento direttamente in produzione.

Durante l'esecuzione di esperimenti di iniezione di guasti, verifica che tutte le parti responsabili ne siano a conoscenza. Comunica ai team appropriati, ad esempio i team responsabili delle operazioni, dell'affidabilità dei servizi e del supporto clienti, quando verranno eseguiti gli esperimenti e l'impatto previsto. Metti a disposizione di questi team strumenti di comunicazione che consentano loro di informare i responsabili dell'esperimento di eventuali effetti avversi.

È necessario ripristinare lo stato originario del carico di lavoro e dei relativi sistemi sottostanti. La progettazione resiliente del carico di lavoro è spesso caratterizzata da funzionalità di riparazione automatica. Tuttavia, alcune progettazioni con difetti o alcuni esperimenti non riusciti possono compromettere in modo imprevisto lo stato del carico di lavoro. Entro la fine dell'esperimento dovrai essere consapevole di questa situazione e ripristinare il carico di lavoro e i sistemi. Con AWS FIS puoi impostare una configurazione di rollback (chiamata anche azione successiva) all'interno dei parametri dell'azione. Una post-operazione ripristina una destinazione allo stato in cui si trovava prima dell'esecuzione dell'operazione stessa. Che siano automatizzate (come l'utilizzo AWS FIS) o manuali, queste azioni relative ai post dovrebbero far parte di un manuale che descrive come rilevare e gestire gli errori.

d. Verifica l'ipotesi.

   Principles of Chaos Engineering fornisce le linee guida su come verificare lo stato stazionario del carico di lavoro.

   È necessario concentrarsi sull'output misurabile di un sistema e non sugli attributi interni del sistema. Le misurazioni di tale output in un breve periodo di tempo costituiscono un'attestazione dello stato stazionario del sistema. Il throughput del sistema nel suo complesso, le percentuali di errori e i percentili della latenza possono essere considerati metriche di interesse che rappresentano il comportamento di uno stato stazionario. Sulla base dei rilevamenti dei modelli di comportamento sistematico durante gli esperimenti, l'ingegneria del caos verifica che il sistema funzioni correttamente anziché tentare di convalidare il modo in cui funziona.

   Nei due esempi precedenti sono state incluse le metriche dello stato stazionario relative a un incremento inferiore allo 0,01% di errori (5xx) lato server e inferiore a un minuto di errori di lettura e scrittura del database.

   Gli errori 5xx rappresentano una buona metrica perché sono la conseguenza della modalità di errore che un client del carico di lavoro sperimenterà direttamente. La misurazione degli errori del database risulta valida come conseguenza diretta dell'errore, ma deve essere supportata da una misurazione diretta dell'impatto, ad esempio le richieste cliente non riuscite o gli errori restituiti a livello di client. Inoltre, includi un monitor sintetico (noto anche come user canary) su qualsiasi dispositivo del tuo carico di lavoro APIs o a cui accede URIs direttamente il client.

e. Migliora la progettazione del carico di lavoro con un occhio di riguardo per la resilienza.

   Se lo stato stazionario non è stato preservato, analizza in che modo puoi migliorare la progettazione del carico di lavoro per azzerare l'impatto dell'errore applicando le best practice

illustrate nel pilastro dell'affidabilità di AWS Well-Architected. Puoi trovare ulteriori informazioni nella AWS Builder's Library, che offre, tra gli altri, articoli su come migliorare i controlli dell'integrità o impiegare nuovi tentativi con backoff nel codice dell'applicazione.

Dopo aver implementato queste modifiche, esegui di nuovo l'esperimento (rappresentato dalla linea punteggiata nel volano relativo all'ingegneria del caos) per determinare la relativa efficacia. Se nella fase di verifica risulta che l'ipotesi è vera, il carico di lavoro sarà in stato stazionario e il ciclo continuerà.

4. Esegui gli esperimenti con regolarità.

   Un esperimento di ingegneria del caos è un ciclo e gli esperimenti devono essere eseguiti regolarmente nell'ambito dell'ingegneria del caos. Se un carico di lavoro è conforme all'ipotesi dell'esperimento, l'esperimento deve essere automatizzato affinché venga eseguito continuamente come fase di regressione della pipeline CI/CD. Per scoprire come eseguire questa operazione, consulta questo blog su come eseguire AWS FIS esperimenti utilizzando. AWS CodePipeline Poi fare pratica con questo lab sugli esperimenti AWS FIS ricorrenti in una pipeline CI/CD.

   Gli esperimenti di iniezione di guasti fanno inoltre parte delle giornate di gioco (consulta REL12-BP06 Conduci regolarmente giornate di gioco). Le giornate di gioco simulano un errore o un evento per verificare sistemi, processi e risposte dei team. Lo scopo è di eseguire effettivamente le azioni che compirebbe il team come se si verificasse un evento eccezionale.

5. Acquisisci e archivia i risultati degli esperimenti.

   I risultati degli esperimenti di iniezione di guasti devono essere acquisiti e resi persistenti. Includi tutti i dati necessari, ad esempio orari, carico di lavoro e condizioni, in modo da essere in grado di analizzare i risultati e i trend in un secondo momento. Esempi di risultati potrebbero includere schermate di dashboard, CSV dump dal database della metrica o un record digitato a mano di eventi e osservazioni dell'esperimento. Puoi inserire la creazione di log degli esperimenti mediante AWS FIS nel processo di acquisizione dei dati.

Risorse

Best practice correlate:

- REL08-BP03 Integra i test di resilienza come parte della tua implementazione
- REL13-BP03 Testare l'implementazione del disaster recovery per convalidare l'implementazione

Documenti correlati:

- Che cos'è? AWS Fault Injection Service
- Che cos'è AWS Resilience Hub?
- Principles of Chaos Engineering
- Chaos Engineering: Planning your first experiment
- Resilience Engineering: Learning to Embrace Failure
- Chaos Engineering stories
- Evitare il fallback nei sistemi distribuiti
- Distribuzione canary per gli esperimenti di ingegneria del caos

Video correlati:

- AWS re:Invent 2020: Test della resilienza utilizzando l'ingegneria del caos () ARC316
- AWS re:Invent 2019: Migliorare la resilienza con l'ingegneria del caos (09-R1) DOP3
- AWS re:Invent 2019: Esecuzione dell'ingegneria del caos in un mondo senza server (01) CMY3

Esempi correlati:

- Well-Architected Lab: Level 300: test per la resilienza di AmazonEC2, Amazon e Amazon S3 RDS
- Chaos Engineering on AWS lab
- Resilient and Well-Architected Apps with Chaos Engineering lab
- Serverless Chaos lab
- Misura e migliora la resilienza delle applicazioni con lab AWS Resilience Hub

Strumenti correlati:

- AWS Fault Injection Service
- Marketplace AWS: Gremlin Chaos Engineering Platform
- Chaos Toolkit
- Chaos Mesh
- Litmus

REL12-BP06 Conduci regolarmente giornate di gioco

Utilizza le giornate di gioco per provare regolarmente le procedure per rispondere a eventi ed errori nel modo più vicino possibile alla produzione (anche negli ambienti di produzione) con le persone che si occuperanno di eventuali scenari di errore reali. Le giornate di gioco applicano misure per garantire che gli eventi di produzione non influiscano sugli utenti.

Le giornate di gioco simulano un errore o un evento per testare sistemi, processi e risposte dei team. Lo scopo è di eseguire effettivamente le azioni che compirebbe il team come se si verificasse un evento eccezionale. Questi ti aiuta a capire dove puoi apportare dei miglioramenti e ti può aiutare a sviluppare un'esperienza organizzativa nella gestione degli eventi. Tali azioni devono essere svolte regolarmente in modo che il team costruisca una "memoria muscolare" su come rispondere.

Quando la progettazione per la resilienza è in loco ed è stata testata in ambienti non di produzione, una giornata di gioco è il modo per garantire che tutto funzioni come pianificato in produzione. Una giornata di gioco, soprattutto la prima, è un'attività di duro lavoro per tutti, in cui tutti gli ingegneri e i team operativi vengono informati in merito a quando accadrà e cosa accadrà. I runbook sono in loco. Gli eventi simulati, compresi i possibili eventi di guasto, vengono eseguiti nei sistemi di produzione nel modo prescritto e ne viene valutato l'impatto. Se tutti i sistemi funzionano come progettato, il rilevamento e la correzione automatica avvengono con un impatto minimo o nullo. Tuttavia, se si osserva un impatto negativo, viene eseguito il rollback del test e i problemi relativi al carico di lavoro vengono risolti, se necessario manualmente (utilizzando il runbook). Poiché le giornate di gioco hanno spesso luogo in produzione, è necessario prendere tutte le precauzioni per garantire che non vi sia alcun impatto sulla disponibilità per i clienti.

Anti-pattern comuni:

- Documentare le procedure senza mai metterle in pratica.
- Non includere i responsabili delle decisioni aziendali negli esercizi di test.

Vantaggi dell'adozione di questa best practice: eseguire giornate di gioco garantisce che tutto il personale segua le policy e le procedure quando si verifica un incidente reale e convalida che tali policy e procedure siano appropriate.

Livello di rischio associato se questa best practice non fosse adottata: medio

Guida all'implementazione

- Programma giornate di gioco per provare regolarmente i tuoi runbook e playbook. Le giornate di gioco devono coinvolgere tutte le persone implicate in un evento di produzione: proprietari di aziende, personale addetto allo sviluppo, personale operativo e team di risposta agli incidenti.
  - Esegui i test di carico o delle prestazioni e successivamente esegui l'iniezione degli errori.
  - Ricerca anomalie nei tuoi runbook e opportunità di provare i tuoi playbook.
    - In caso di deviazione dai tuoi runbook, perfeziona il runbook o correggi il comportamento. Se ti eserciti sul tuo playbook, identifica il runbook che avrebbe dovuto essere usato, oppure creane uno nuovo.

Risorse

Video correlati:

- AWS re:Invent 2019: Migliorare la resilienza con l'ingegneria del caos (09-R1) DOP3

  Esempi correlati:

- AWS Well-Architected Labs: test di resilienza

REL13. Come si pianifica il ripristino di emergenza?

Avere backup e componenti del carico di lavoro ridondanti in loco è l'inizio della strategia di ripristino di emergenza. RTOe RPO sono i vostri obiettivi per il ripristino del carico di lavoro. Imposta questi valori in base alle esigenze aziendali. Implementa una strategia per raggiungere questi obiettivi, prendendo in considerazione le posizioni e la funzione delle risorse e dei dati del carico di lavoro. La probabilità di interruzione e il costo del ripristino sono fattori chiave che aiutano a comunicare il valore aziendale che può avere il ripristino di emergenza per un carico di lavoro.

Best practice

- REL13-BP01 Definire gli obiettivi di ripristino per tempi di inattività e perdita di dati
- REL13-BP02 Utilizzare strategie di ripristino definite per raggiungere gli obiettivi di ripristino
- REL13-BP03 Testare l'implementazione del disaster recovery per convalidare l'implementazione
- REL13-BP04 Gestire la deriva della configurazione nel sito o nella regione di DR
- REL13-BP05 Ripristino automatico

REL13-BP01 Definire gli obiettivi di ripristino per tempi di inattività e perdita di dati

Il carico di lavoro ha un obiettivo per il tempo di ripristino (RTO) e un obiettivo per il punto di ripristino (RPO).

Recovery Time Objective (RTO) è il ritardo massimo accettabile tra l'interruzione del servizio e il ripristino del servizio. Questo determina ciò che viene considerato un intervallo di tempo accettabile in caso di indisponibilità del servizio.

Recovery Point Objective (RPO) è 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.

RTOe RPO i valori sono considerazioni importanti nella scelta di una strategia di Disaster Recovery (DR) appropriata per il carico di lavoro. Tali obiettivi sono stabiliti dall'azienda e poi vengono utilizzati dai team tecnici per selezionare e implementare una strategia di ripristino di emergenza.

Risultato desiderato:

A ogni carico di lavoro è assegnato RTO e RPO definito in base all'impatto aziendale. Il carico di lavoro viene assegnato a un livello predefinito, che definisce la disponibilità del servizio e la perdita accettabile di dati, con un livello associato e. RTO RPO Se tale livello non è raggiungibile, è possibile assegnare un livello personalizzato per carico di lavoro, con l'obiettivo di creare i livelli in un secondo momento. RTOe RPO vengono utilizzati come una delle considerazioni principali per la selezione di un'implementazione della strategia di disaster recovery per il carico di lavoro. Altre riflessioni nel momento della scelta di una strategia di ripristino di emergenza sono i vincoli economici, le dipendenze del carico di lavoro e i requisiti operativi.

InfattiRTO, comprendi l'impatto in base alla durata di un'interruzione. È lineare o ci sono implicazioni non lineari (ad esempio, dopo 4 ore, chiudi una linea di produzione fino l'inizio del turno successivo)?

Una matrice di ripristino di emergenza, come quella seguente, può aiutarti a capire come la criticità del carico di lavoro sia collegata agli obiettivi di ripristino (da notare che i valori reali per gli assi X e Y devono essere personalizzati in base alle esigenze della tua organizzazione).

Matrice di ripristino di emergenza						
		Obiettivo del punto di ripristino				
		meno di 1 minuto	meno di 1 ora	meno di 6 ore	meno di 1 giorno	Più di 1 giorno
Obiettivo del tempo di ripristino	meno di 10 minuti	Critica	Critica	Alta	Medio	Medio
	meno di 2 ore	Critica	Alta	Medio	Medio	Bassa
	meno di 8 ore	Alta	Medio	Medio	Bassa	Bassa
	meno di 24 ore	Medio	Medio	Bassa	Bassa	Bassa
	Più di 24 ore	Medio	Bassa	Bassa	Bassa	Bassa

Figura 16: matrice di ripristino di emergenza

Anti-pattern comuni:

- Nessun obiettivo di ripristino definito.
- Selezione di obiettivi di ripristino arbitrari.
- Selezione di obiettivi di ripristino troppo tolleranti e che non soddisfano gli obiettivi aziendali.
- Mancanza di comprensione dell'impatto dei tempi di inattività e perdita dei dati.
- Selezione di obiettivi di ripristino non realistici, come tempo zero di ripristino e nessuna perdita di dati, che potrebbero non essere raggiungibili per la configurazione del tuo carico di lavoro.
- Selezione di obiettivi di ripristino più severi rispetto agli obiettivi aziendali effettivi. Questo costringe a effettuare implementazioni di ripristino di emergenza più costose e complicate rispetto alle esigenze del carico di lavoro.
- Selezione di obiettivi di ripristino non compatibili con quelli di un carico di lavoro dipendente.
- I tuoi obiettivi di ripristino non considerano i requisiti di conformità normativa.
- RTOe RPO definito per un carico di lavoro, ma mai testato.

Vantaggi dell'adozione di questa best practice: gli obiettivi di ripristino in termini di tempo e perdita di dati sono necessari per guidare l'implementazione del ripristino di emergenza.

Livello di rischio associato se questa best practice non fosse adottata: elevato

Guida all'implementazione

Per un dato carico di lavoro devi considerare l'impatto dei tempi di inattività e della perdita dei dati per la tua azienda. L'impatto generalmente aumenta all'aumentare dei tempi di inattività o della perdita dei dati, ma il ritmo di tale crescita cambia in base al tipo di carico di lavoro. Ad esempio, potresti tollerare l'inattività per massimo un'ora con conseguenze minime, ma successivamente l'impatto diventerebbe rapidamente più serio. L'impatto sull'azienda si manifesta in forme diverse, tra cui costi economici (come perdita di fatturato), fiducia del cliente (e impatto sulla reputazione), problematiche operative (come stipendi in ritardo o diminuzione della produttività) e rischi normativi. Utilizza i passaggi seguenti per comprendere questi impatti e impostare RTO e RPO per il tuo carico di lavoro.

Passaggi dell'implementazione

1. Individua le parti interessate aziendali per questo carico di lavoro e collabora con loro per implementare questi passaggi. Gli obiettivi di ripristino di un carico di lavoro sono il frutto di una decisione aziendale. I team tecnici, quindi, lavorano con le parti interessate aziendali e usano questi obiettivi per selezionare una strategia di ripristino di emergenza.

   > **Note**
   >
   > Per i passaggi 2 e 3 puoi usare il the section called “Foglio di lavoro di implementazione”.

2. Raccogli le informazioni necessarie per prendere una decisione rispondendo alle domande qui di seguito.
3. Hai categorie o livelli di criticità in termini di impatto del tuo carico di lavoro nella tua organizzazione?
   a. Se sì, assegna questo carico di lavoro a una categoria
   b. Se no, definisci queste categorie. Crea al massimo cinque categorie e perfeziona l'intervallo del tuo obiettivo del tempo di ripristino (RTO) per ognuna. Ecco alcune categorie di esempio: critico, alto, medio, basso. Per capire come mappare i carichi di lavoro rispetto alle categorie devi considerare se il carico di lavoro è mission-critical, importante per l'azienda o non trainante.
   c. Imposta il carico di lavoro RTO e RPO in base alla categoria. Scegli sempre una categoria più rigorosa (inferiore RTO eRPO) rispetto ai valori grezzi calcolati in questo passaggio. Se ciò comporta una variazione significativa di valore non rispondente alle esigenze, prendi in considerazione la possibilità di creare una nuova categoria.

4. In base a queste risposte, RTO assegna dei RPO valori al carico di lavoro. Puoi farlo direttamente o assegnando il carico di lavoro a un livello predefinito di servizio.
5. Documenta il piano di disaster recovery (DRP) per questo carico di lavoro, che fa parte del piano di continuità aziendale dell'organizzazione (BCP), in una posizione accessibile al team addetto al carico di lavoro e alle parti interessate
   a. Registrate la RTO e RPO le informazioni utilizzate per determinare questi valori. Includi la strategia utilizzata per valutare l'impatto del carico di lavoro sull'azienda.
   b. Registra anche altre metriche RTO e RPO stai monitorando o hai intenzione di monitorare gli obiettivi di disaster recovery
   c. Dopo aver creato questi valori, potrai aggiungere i dettagli della tua strategia di ripristino di emergenza e il runbook.
6. Osservando le criticità del carico di lavoro in una matrice come quella della Figura 15, puoi iniziare a stabilire livelli predefiniti di servizio per la tua organizzazione.
7. Dopo aver implementato una strategia di DR (o una dimostrazione di fattibilità di una strategia di DR) come indicato in precedenzathe section called “REL13-BP02 Utilizzare strategie di ripristino definite per raggiungere gli obiettivi di ripristino”, testala per determinare il carico di lavoro effettivo RTC (Recovery Time Capability) e RPC (Recovery Point Capability). Se questi valori non sono in linea con gli obiettivi target di ripristino, puoi collaborare con le parti interessate della tua azienda per modificarli o cambiare la strategia di ripristino di emergenza in modo che possa soddisfare tali obiettivi.

Domande principali

1. Qual è il tempo massimo durante il quale il carico di lavoro può essere inattivo prima che questo abbia un impatto grave sull'attività?
   a. Definisci il costo monetario (impatto finanziario diretto) sull'attività al minuto se il carico di lavoro è inattivo.
   b. Considera che l'impatto non è sempre lineare. L'impatto può essere limitato all'inizio e poi aumentare rapidamente oltre un punto critico specifico.
2. Qual è la quantità massima di dati che possiamo perdere prima che questo abbia un impatto grave sull'attività?
   a. Considera questo valore per gli archivi di dati più strategici. identifica le criticità relative ad altri archivi di dati.

   b. I dati del carico di lavoro possono essere ricreati se persi? Se dal punto di vista operativo è più semplice del backup e del ripristino, scegli in RPO base alla criticità dei dati di origine utilizzati per ricreare i dati del carico di lavoro.

3. Quali sono gli obiettivi di ripristino e le aspettative di disponibilità dei carichi di lavoro da cui questo valore dipende (a valle) o i carichi di lavoro che dipendono da questo valore (a monte)?
   a. Scegli obiettivi di ripristino che consentono a questo carico di lavoro di soddisfare i requisiti delle dipendenze a monte.
   b. Scegli obiettivi di ripristino che sono raggiungibili considerate le funzionalità di ripristino delle dipendenze a valle. Possono essere escluse le dipendenze downstream non critiche (quelle che puoi "aggirare"). In alternativa, lavora con dipendenze downstream critiche per migliorare le funzionalità di ripristino, laddove necessario.

Domande aggiuntive

Considera queste domande e come possono essere applicate a questo carico di lavoro:

4. Le interruzioni sono diverse RTO e RPO dipendono dal tipo di interruzione (regione o zona, ecc.)?
5. C'è un momento specifico (stagionalità, eventi di vendita, lanci di prodotti) in cui la tua/potrebbe cambiare? RTO RPO Se sì, qual è la misurazione diversa e il vincolo temporale?
6. Se il carico di lavoro viene interrotto, quanti clienti ne subiranno l'impatto?
7. Qual è l'impatto sulla reputazione se il carico di lavoro viene interrotto?
8. Quali altri impatti operativi possono verificarsi se il carico di lavoro viene interrotto? Ad esempio, l'impatto sulla produttività dei dipendenti se i sistemi e-mail non sono disponibili o sei i sistemi di buste paga non sono in grado di inviare le transazioni.
9. In che modo vengono caricati i carichi di lavoro RTO e in che modo si RPO allineano alla strategia di disaster recovery aziendale e organizzativa?
10. Esistono obblighi contrattuali interni per fornire un servizio? Esistono delle penali nel caso in cui non siano soddisfatti?
11. Quali sono i limiti normativi o di conformità dei dati?

Foglio di lavoro di implementazione

Puoi usare questo foglio di lavoro per le fasi 2 e 3 dell'implementazione. Adegua questo foglio di lavoro in base alle tue esigenze specifiche, aggiungendo, ad esempio, altre domande.

Passo 2: domande principali	Si applica al carico di lavoro?	RTO del carico di lavoro	RPO del carico di lavoro	RTO rettif.	RPO rettif.	Istruzioni
[1] tempo massimo di inattività del carico di lavoro						misurato in tempo dall'inizio del malfunzionamento al ripristino
[2] quantità massima di dati che possono essere persi						misurato in tempo trascorso dall'ultimo set di dati integro ripristinabile
[3a] dipendenze a monte						inserire gli obiettivi di recupero a monte più rigorosi
[3b] riconciliazione delle dipendenze a valle						inserire gli obiettivi di recupero a valle meno rigorosi
[3a] riconciliazione delle dipendenze a monte						Se il valore a monte è inferiore ai valori attuali e il valore a valle è superiore,
[3b] riconciliazione delle dipendenze a valle						operare sulle dipendenze per riconciliare i valori e inserirli qui.
[3] dipendenze						ridurre i valori per soddisfare le dipendenze a monte o alzarli in base alle capacità delle dipendenza a valle
Passo 2: domande aggiuntive						Indicare se la domanda è pertinente. Saltarla in caso affermativo
RTO/RPO di base						Riportare qui i valori di RTO e RPO sopra indicati
[4] tipo di malfunzionamento	[ ]S / [ ]N					Inserire gli obiettivi di recupero per i tipi di evento con i requisiti più rigorosi
[5] obiettivi specifici basati sul tempo	[ ]S / [ ]N					Inserire gli obiettivi di recupero per i tempi con i requisiti più rigorosi
[6] clienti che sperimentano il disservizio	[ ]S / [ ]N					Tracciare un grafico dei clienti che sperimentano il disservizio in funzione del tempo di inattività o dei dati persi. Utilizzare tale grafico per inserire i valori massimi di RTO e RPO ammissibili in base all'impatto sui clienti
[7] impatto reputazionale	[ ]S / [ ]N					Lavorare in modo congiunto con l'azienda per determinare i massimi valori di RTO e RPO in base all'impatto sulla reputazione
[8] impatto operativo	[ ]S / [ ]N					Inserire i valori massimi di RTO e RPO sulla base dell'impatto operativo
[9] allineamento aziendale	[ ]S / [ ]N					Inserire i valori massimi di RTO e RPO per i carichi di lavoro di questo tipo in base ai requisiti LOB e organizzativi
[10] obblighi contrattuali	[ ]S / [ ]N					Inserire i valori massimi di RTO e RPO sulla base degli obblighi contrattuali
[11] conformità normativa	[ ]S / [ ]N					Inserire i valori massimi di RTO e RPO sulla base delle norme di conformità applicabili
obiettivo sulla base delle domande aggiuntive						Selezionare il valore minimo (valore più rigoroso) dalle domande 4-11 e inserirlo qui
obiettivo rettificato						Se non è possibile raggiungere gli obiettivi indicati nella riga precedente, collaborare con le parti interessate per allentare i vincoli e inserire un nuovo minimo qui.
RTO/RPO rettificato						Inserire il valore inferiore tra RPO/RTO di base e valore obiettivo rettificato
Passo 3						
Mappatura su categorie o livelli predefiniti						Regolare entrambi i valori verso il basso (requisito più rigoroso) per allinearsi al livello più vicino definito

Foglio di lavoro

Livello di impegno per il piano di implementazione: basso

Risorse

Best practice correlate:

- the section called “REL09-BP04 Eseguire il ripristino periodico dei dati per verificare l'integrità e i processi di backup”
- the section called “REL13-BP02 Utilizzare strategie di ripristino definite per raggiungere gli obiettivi di ripristino”
- the section called “REL13-BP03 Testare l'implementazione del disaster recovery per convalidare l'implementazione”

Documenti correlati:

- AWS Architecture Blog: serie sul ripristino di emergenza
- Disaster recovery dei carichi di lavoro su AWS: Recovery in the Cloud (white paper)AWS

- Gestione delle politiche di resilienza con Resilience Hub AWS
- APNPartner: partner che possono contribuire al disaster recovery
- Marketplace AWS: prodotti utilizzabili per il ripristino di emergenza

Video correlati:

- AWS re:Invent 2018: modelli di architettura per applicazioni Active-Active in più regioni (09-R2) ARC2
- Ripristino di emergenza dei carichi di lavoro su AWS

REL13-BP02 Utilizzare strategie di ripristino definite per raggiungere gli obiettivi di ripristino

Definisci una strategia di ripristino di emergenza (DR) che soddisfi gli obiettivi di ripristino del carico di lavoro. Scegli una strategia, ad esempio backup e ripristino, standby (attivo/passivo) o attivo/attivo.

Risultato desiderato: per ciascun carico di lavoro esiste una strategia di ripristino di emergenza definita e implementata che consente a quel carico di lavoro di raggiungere gli obiettivi di ripristino. Le strategie di ripristino di emergenza tra carichi di lavoro utilizzano modelli riutilizzabili (come strategie descritte in precedenza),

Anti-pattern comuni:

- Implementazione di procedure di ripristino incoerenti per carichi di lavoro con obiettivi di ripristino simili.
- Implementazione di una strategia di ripristino di emergenza ad-hoc quando si verifica un disastro.
- Assenza di piani per il ripristino di emergenza.
- Dipendenza dalle operazioni del piano di controllo (control-plane) durante il ripristino.

Vantaggi dell'adozione di questa best practice:

- L'utilizzo di strategie di ripristino definite consente di utilizzare strumenti e procedure di test comuni.
- L'uso di strategie di ripristino definite permette la condivisione delle informazioni tra team e l'implementazione del ripristino di emergenza nei carichi dl lavoro di loro proprietà.

Livello di rischio associato se questa best practice non fosse adottata: elevato. Senza una strategia di ripristino di emergenza pianificata, implementata e testata, è poco probabile riuscire a raggiungere gli obiettivi di ripristino in caso di emergenze.

Guida all'implementazione

Una strategia di ripristino di emergenza si basa sulla capacità di creare il tuo carico di lavoro in un sito di ripristino se la tua sede principale non è disponibile per eseguire il carico di lavoro. Gli obiettivi di ripristino più comuni sono RTO eRPO, come illustrato in. REL13-BP01 Definire gli obiettivi di ripristino per tempi di inattività e perdita di dati

Una strategia di ripristino di emergenza su più zone di disponibilità (AZs) all'interno di un'unica Regione AWS zona può fornire una mitigazione in caso di eventi di emergenza come incendi, inondazioni e gravi interruzioni di corrente. Se è necessario implementare la protezione da un evento improbabile che impedisce l'esecuzione del carico di lavoro in una determinata area Regione AWS, è possibile utilizzare una strategia di disaster recovery che utilizzi più regioni.

Quando pianifichi una strategia di ripristino di emergenza su più regioni, devi scegliere una delle seguenti strategie. Sono elencate in ordine crescente di costo e complessità e in ordine decrescente di RTO e. RPO La regione di ripristino si riferisce a una regione Regione AWS diversa da quella principale utilizzata per il carico di lavoro.

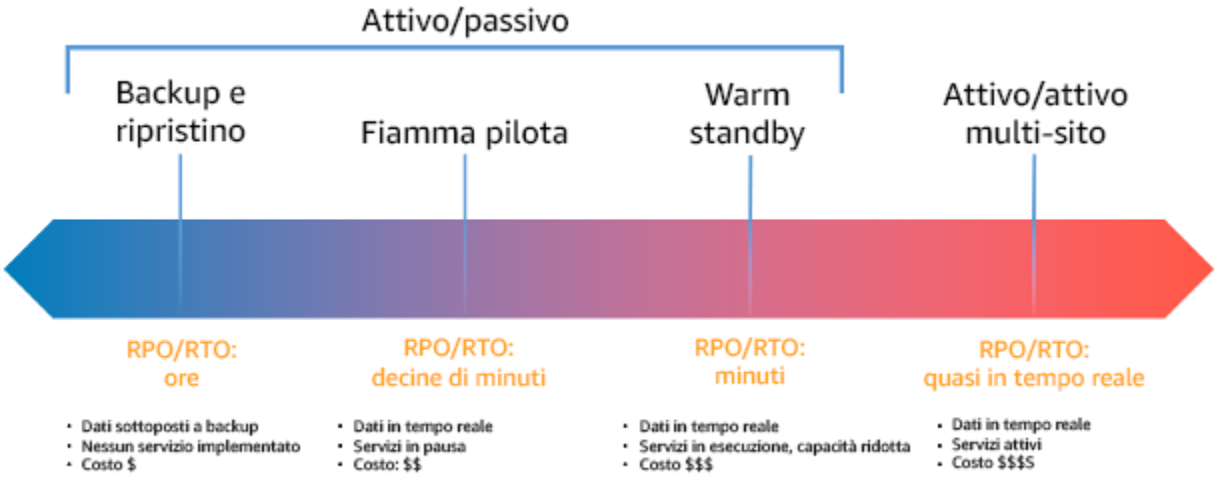


Figura 17: strategie di ripristino di emergenza (DR)

- Backup e ripristino (RPOin poche ore, RTO in 24 ore o meno): esegui il backup dei dati e delle applicazioni nell'area di ripristino. L'utilizzo di backup automatici o continui consentirà il ripristino point-in-time (PITR), che in alcuni casi può ridursi RPO fino a 5 minuti. In caso di emergenza, sarà necessario implementare l'infrastruttura (utilizzando l'infrastruttura come codice per ridurreRTO), distribuire il codice e ripristinare i dati di backup per il ripristino in caso di emergenza nella regione di ripristino.
- Pilot light (RPOin pochi minuti, RTO in decine di minuti): esegui il provisioning di una copia dell'infrastruttura di carico di lavoro principale nella regione di ripristino. Replica i dati nella regione di ripristino e crea un backup in essa. Le risorse necessarie per supportare la replica dei dati e il backup, come database e archiviazione di oggetti, sono sempre attive. Altri elementi come i server applicativi o il calcolo serverless non vengono distribuiti, ma possono essere creati quando necessari con la configurazione e il codice applicativo richiesti.
- Warm standby (RPOin secondi, RTO in minuti): mantieni una versione ridotta ma perfettamente funzionante del tuo carico di lavoro sempre in esecuzione nella regione di ripristino. I sistemi business critical sono completamente duplicati e sono sempre accesi, ma con un parco istanze ridotto verticalmente. I dati vengono replicati e si trovano nella regione di recupero. Al momento del ripristino, il sistema viene fatto aumentare verticalmente rapidamente per gestire il carico di produzione. Quanto più è elevato lo standby a caldo, tanto minore RTO sarà il ricorso al piano di controllo. Quando il dimensionamento è completo, si parla di standby a caldo.
- Multiregione (multisito) attivo-attivo (RPOvicino allo zero, RTO potenzialmente zero): il carico di lavoro viene distribuito e servito attivamente dal traffico proveniente da più regioni. Regioni AWS Questa strategia comporta la sincronizzazione dei dati tra le regioni. È necessario evitare o gestire possibili conflitti causati da scritture sullo stesso record in due diverse repliche regionali, un'attività che potrebbe rivelarsi complessa. La replica dei dati è utile per la sincronizzazione dei dati e vi proteggerà da alcuni tipi di emergenza, ma non dal danneggiamento o dalla distruzione dei dati a meno che la soluzione non includa anche opzioni di ripristino. point-in-time

> **Note**
>
> La differenza tra Pilot Light e Warm Standby può talvolta essere difficile da comprendere. Entrambe prevedono un ambiente nella tua regione di ripristino con copie degli asset della tua regione principale. La differenza è che la strategia Pilot Light non può elaborare le richieste senza aver prima intrapreso altre azioni, mentre Warm Standby può gestire immediatamente il traffico (a livelli ridotti di capacità). La strategia Pilot Light richiede l'attivazione dei server, possibilmente l'implementazione di un'infrastruttura aggiuntiva (non principale) e l'aumentare verticalmente, mentre Warm Standby richiede solo l'aumentare

verticalmente (tutto è già stato implementato ed è in esecuzione). Scegliete tra queste opzioni in base alle vostre esigenzeRTO. RPO
Quando il costo è un problema e desideri raggiungere RTO obiettivi simili RPO a quelli definiti nella strategia warm standby, potresti prendere in considerazione soluzioni native per il cloud AWS Elastic Disaster Recovery, come quelle che adottano l'approccio «pilot light» e offrono RTO obiettivi miglioratiRPO.

Passaggi dell'implementazione

1. Definisci una strategia di ripristino di emergenza in linea con i requisiti di ripristino di questo carico di lavoro.

   La scelta di una strategia di disaster recovery è un compromesso tra la riduzione dei tempi di inattività e della perdita di dati (RTOeRPO) e il costo e la complessità dell'implementazione della strategia. Dovresti evitare di implementare una strategia che sia più severa del necessario, in quanto questo comporterebbe costi aggiuntivi.

   Ad esempio, nel diagramma seguente, l'azienda ha determinato il massimale consentito e il limite di RTO quanto può spendere per la propria strategia di ripristino dei servizi. Considerati gli obiettivi aziendali, le strategie DR Pilot Light o Warm Standby soddisferanno sia i criteri di costo che quelli relativi ai RTO costi.

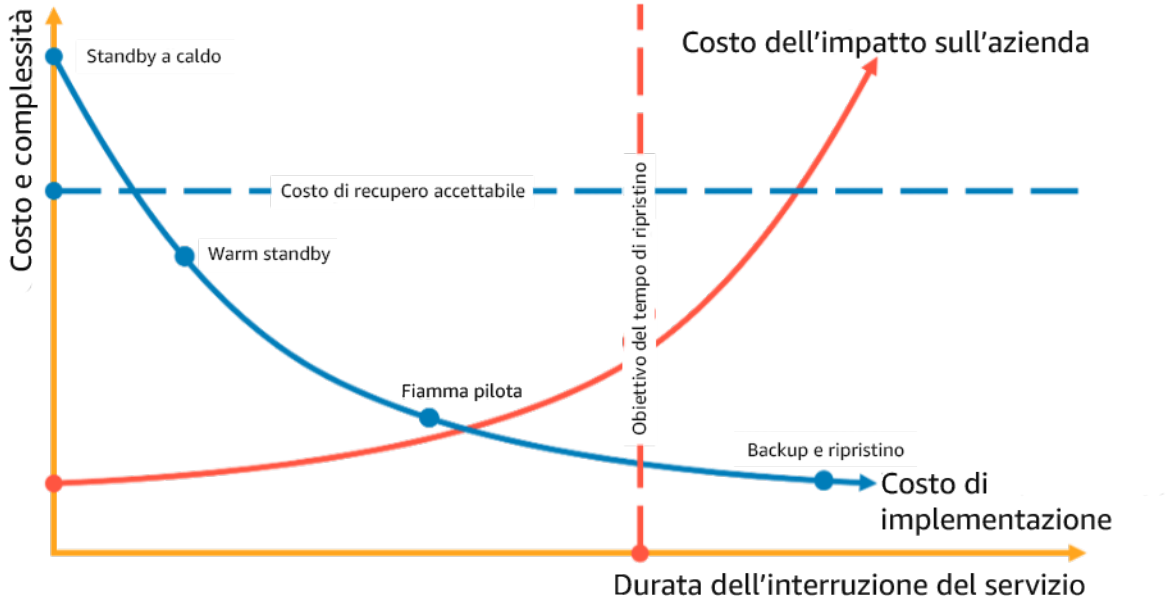


Figura 18: Scelta di una strategia di DR basata sui RTO costi

Per ulteriori informazioni, vedere Business Continuity Plan (BCP).

2. Esamina i modelli con cui la strategia di ripristino di emergenza selezionata può essere implementata.

   Questo passaggio consiste nel capire come implementare la strategia selezionata. Le strategie vengono spiegate utilizzando Regioni AWS come siti primari e di ripristino. Tuttavia, puoi anche decidere di utilizzare le zone di disponibilità in una singola regione come strategia di ripristino di emergenza, utilizzando aspetti di più strategie.

   Nei passaggi seguenti puoi applicare la strategia al carico di lavoro specifico.

   Backup e ripristino

   Il backup e il ripristino sono la strategia meno complessa da implementare, ma richiederanno più tempo e impegno per ripristinare il carico di lavoro, con conseguente aumento RTO della capacità di risposta. RPO È buona norma eseguire sempre backup dei dati e copiarli su un altro sito (ad esempio un altro Regione AWS).

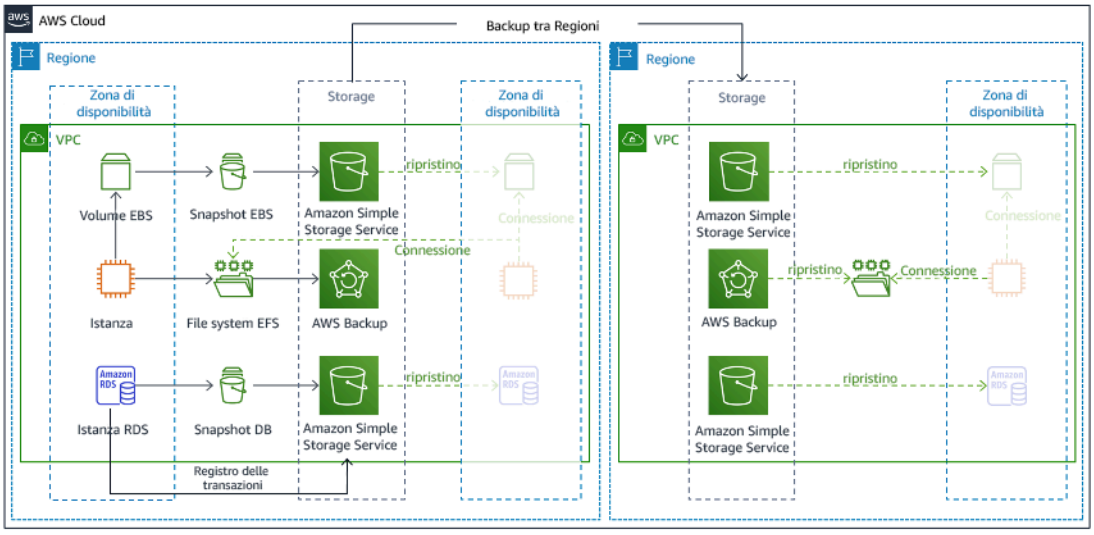


Figura 19: architettura di backup e ripristino

Per ulteriori dettagli su questa strategia, vedere Disaster Recovery (DR) Architecture on AWS, Parte II: Backup e ripristino con Rapid Recovery.

Pilot light

Con l'approccio pilot light, replichi i dati dalla tua regione principale alla regione di ripristino. Le risorse di base utilizzate per l'infrastruttura del carico di lavoro vengono implementate nella regione di ripristino; tuttavia sono comunque necessarie risorse aggiuntive ed eventuali dipendenze per rendere funzionale questo stack. Ad esempio, nella Figura 20 non viene implementata alcuna risorsa di calcolo.

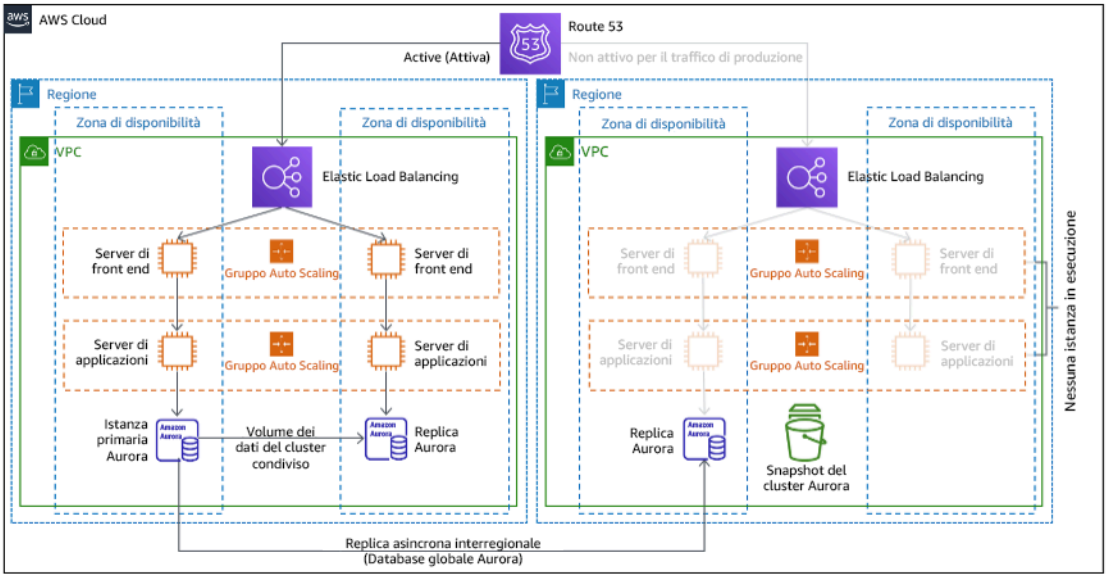


Figura 20: architettura pilot light

Per ulteriori dettagli su questa strategia, vedere Disaster Recovery (DR) Architecture on AWS, ParteIII: Pilot Light and Warm Standby.

Warm standby

L'approccio warm standby implica la verifica della presenza di una copia ridotta verticalmente, ma comunque funzionale, dell'ambiente di produzione in un'altra regione. Questo approccio estende il concetto di Pilot Light e diminuisce il tempo di ripristino, poiché il carico di lavoro è sempre attivo in un'altra regione. Se la regione di ripristino ha raggiunto il massimo della capacità, allora viene definita come standby a caldo.

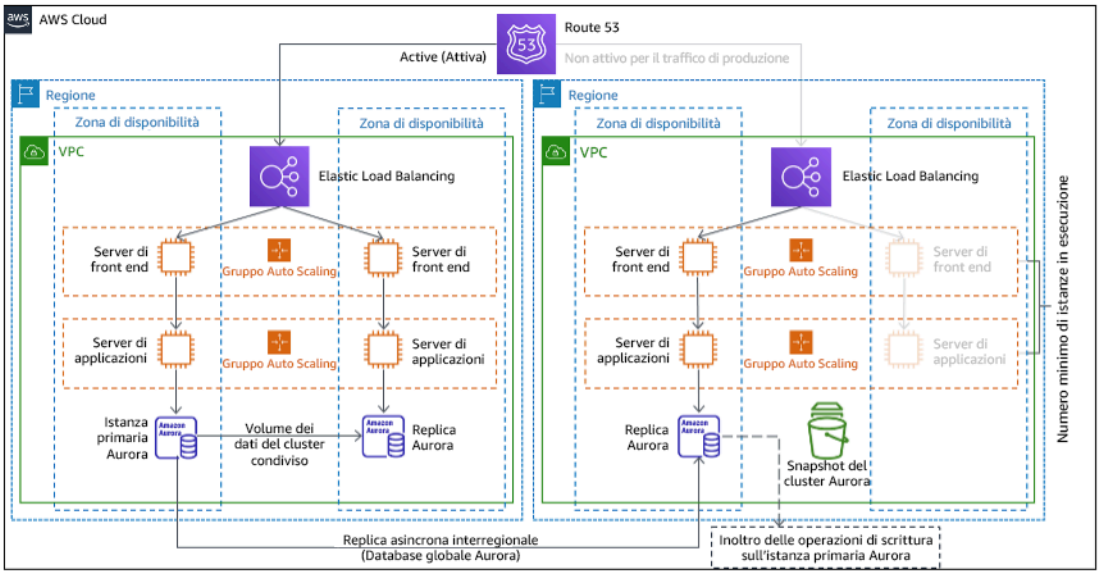


Figura 21: architettura warm standby

Se si utilizza Warm Standby o Pilot Light è necessario aumentare verticalmente le risorse nella regione di ripristino. Per verificare che la capacità sia disponibile quando necessario, prendi in considerazione l'utilizzo per le prenotazioni di capacità per le EC2 istanze. In caso di utilizzo AWS Lambda, la concorrenza con provisioning può fornire ambienti di runtime in modo che siano pronti a rispondere immediatamente alle chiamate della funzione.

Per maggiori dettagli su questa strategia, consulta Disaster Recovery (DR) Architecture on AWS, PartIII: Pilot Light and Warm Standby.

Attivo/attivo multi-sito

Puoi eseguire il carico di lavoro simultaneamente in più regioni come parte di una strategia attivo/attivo multi-sito. La strategia attivo/attivo multi-sito serve il traffico da tutte le regioni in cui è distribuita. I clienti possono selezionare questa strategia per motivi diversi dal ripristino di emergenza. Può essere utilizzata per aumentare la disponibilità o nella distribuzione di un carico di lavoro a un pubblico globale (per posizionare l'endpoint più vicino agli utenti e/o per distribuire stack localizzati al pubblico di quella regione). Come strategia di disaster recovery, se il carico di lavoro non può essere supportato in una delle Regioni AWS aree in cui è distribuito, tale regione viene evacuata e le regioni rimanenti vengono utilizzate per mantenere la disponibilità. La strategia attivo/attivo multi-sito è la strategia di ripristino operativamente più complessa e dovrebbe essere selezionata solo quando lo richiedono i requisiti aziendali.

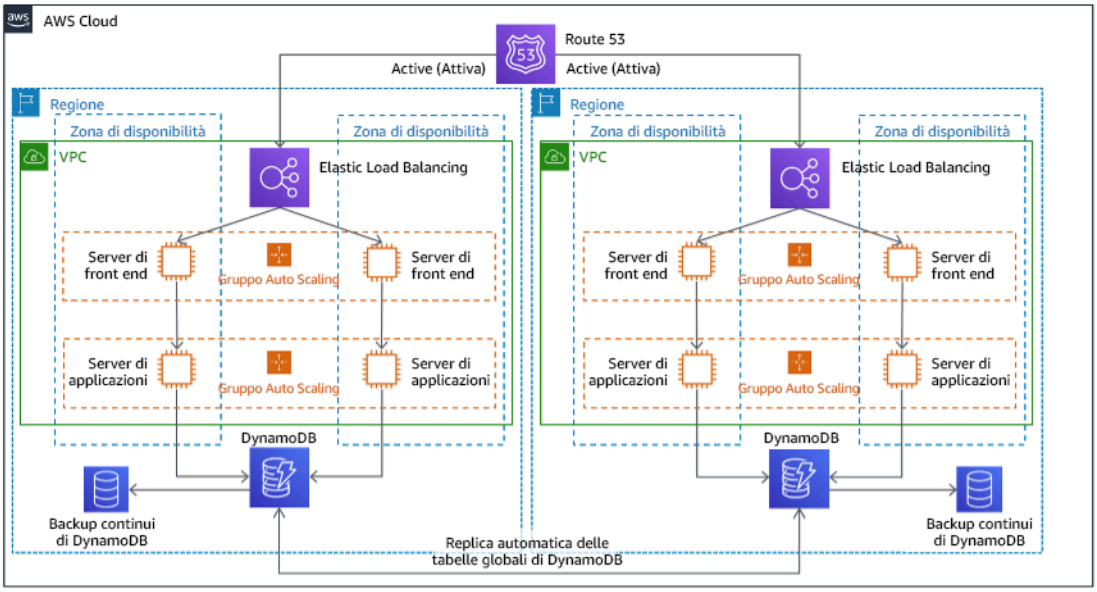


Figura 22: architettura attivo/attivo multi-sito

Per ulteriori dettagli su questa strategia, vedere Disaster Recovery (DR) Architecture on AWS, Parte IV: Multi-site Active/Active.

AWS Elastic Disaster Recovery

Se state prendendo in considerazione la strategia Pilot Light o Warm Standby per il disaster recovery, AWS Elastic Disaster Recovery potrebbe fornire un approccio alternativo con maggiori vantaggi. Elastic Disaster Recovery può offrire un RPO RTO obiettivo simile a quello dello standby a caldo, ma mantiene l'approccio a basso costo della luce pilota. Elastic Disaster Recovery replica i dati dalla regione principale alla regione di ripristino, utilizzando una protezione continua dei dati per ottenere un risultato RPO misurabile in secondi e un RTO valore misurabile in minuti. Solo le risorse necessarie per replicare i dati vengono implementate nella regione di ripristino, mantenendo i costi ridotti come nella strategia Pilot Light. Quando usi Elastic Disaster Recovery, il servizio coordina e orchestra il ripristino delle risorse di calcolo quando viene avviato come parte di un failover o di un'esercitazione.

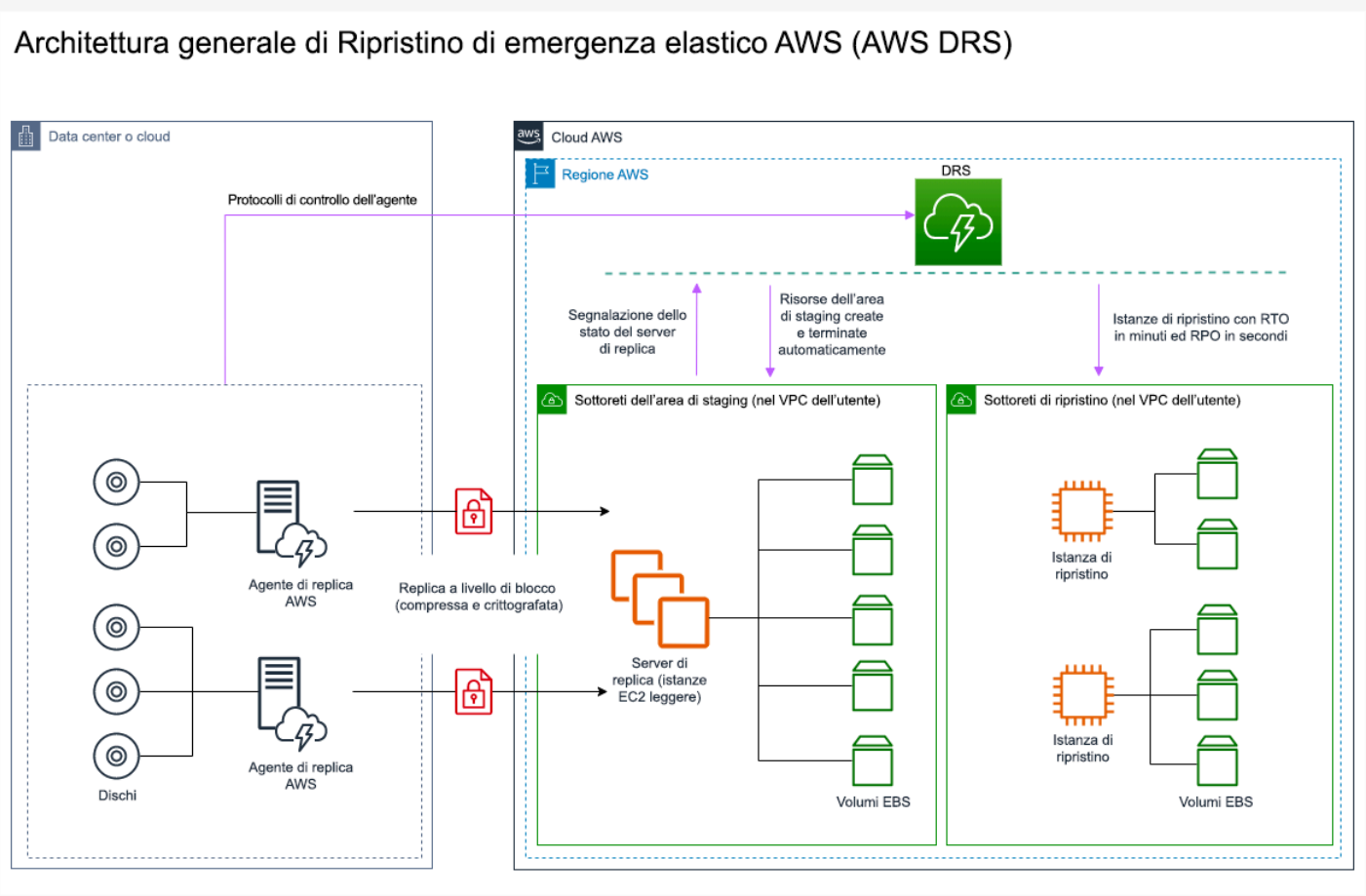


Figura 23: architettura AWS Elastic Disaster Recovery

Procedure aggiuntive per la protezione dei dati

Con tutte le strategie devi anche mitigare un disastro relativo ai dati. La replica continua dei dati protegge da alcuni tipi di emergenza, ma potrebbe non proteggervi dal danneggiamento o dalla distruzione dei dati, a meno che la strategia non includa anche il controllo delle versioni dei dati archiviati o opzioni di point-in-time ripristino. È inoltre necessario eseguire il backup dei dati replicati nel sito di ripristino per creare point-in-time backup oltre alle repliche.

Utilizzo di più zone di disponibilità (AZs) all'interno di una singola Regione AWS

Quando ne vengono utilizzate più di una AZs all'interno di una singola regione, l'implementazione del DR utilizza più elementi delle strategie di cui sopra. Innanzitutto è necessario creare un'architettura ad alta disponibilità (HA), utilizzandone più di una, AZs come illustrato nella Figura 23. Questa architettura utilizza un approccio attivo/attivo multisito, poiché EC2le istanze Amazon e Elastic Load Balancer dispongono di risorse distribuite in più richieste che gestiscono attivamente.

AZs L'architettura dimostra anche l'hot standby, in cui se l'RDSistanza Amazon primaria si guasta (o la stessa AZ fallisce), l'istanza in standby viene promossa a primaria.

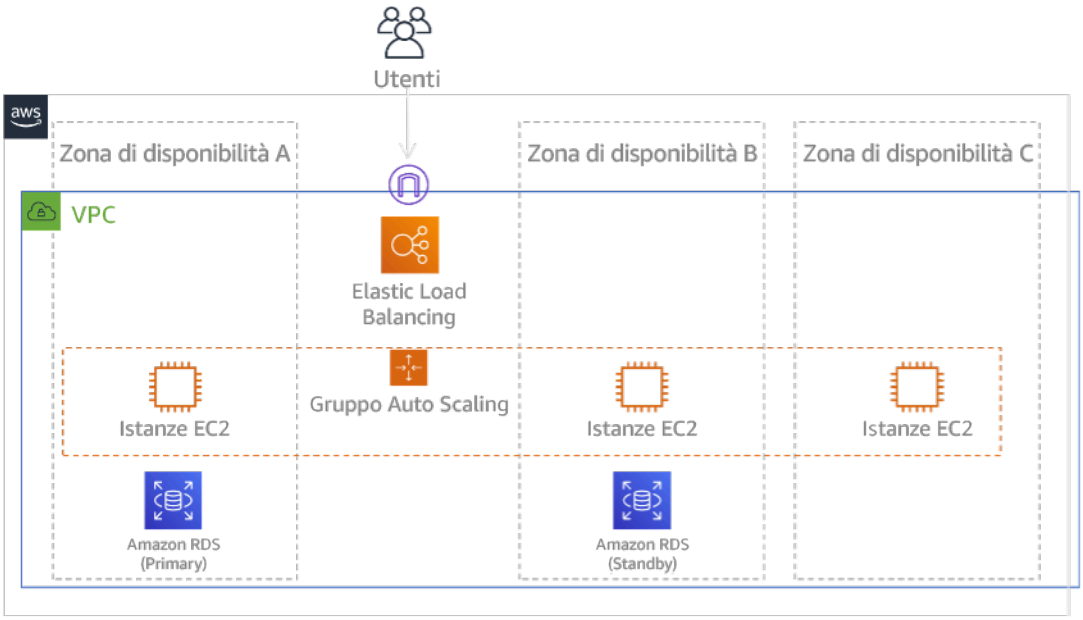


Figura 24: architettura con più zone di disponibilità

Oltre a questa architettura HA, devi aggiungere i backup di tutti i dati richiesti per eseguire il tuo carico di lavoro. Ciò è particolarmente importante per i dati vincolati a una singola zona, come i EBSvolumi Amazon o i cluster Amazon Redshift. In caso di errore di una zona di disponibilità, dovrai ripristinare i dati in un'altra zona di disponibilità. Ove possibile, dovresti anche copiare i backup dei dati su un altro Regione AWS come ulteriore livello di protezione.

Un approccio alternativo meno comune alla singola regione, Multi-AZ DR è illustrato nel post di blog, Creazione di applicazioni altamente resilienti utilizzando Amazon Application Recovery Controller, Parte 1: stack per regione singola. In questo caso, la strategia consiste nel mantenere il maggior isolamento possibile tra le regioni, ad AZs esempio il modo in cui operano le regioni. Usando questa strategia alternativa puoi scegliere un approccio attivo/attivo o attivo/passivo.

> **Note**
>
> Alcuni carichi di lavoro hanno requisiti normativi di residenza dei dati. Se ciò si applica al carico di lavoro in una località che attualmente ne ha una sola Regione AWS, Multiregione non è la soluzione ideale per le esigenze aziendali. Le strategie con più zone di disponibilità offrono una buona protezione dalla maggior parte dei disastri.

3. Valuta le risorse del tuo carico di lavoro e quale sarà la loro configurazione nella regione di ripristino prima del failover (durante la normale operatività).

   Per l'infrastruttura e AWS le risorse, utilizza l'infrastruttura come codice AWS CloudFormationo strumenti di terze parti come Hashicorp Terraform. Puoi utilizzare la distribuzione su più account e regioni con un'unica operazione. AWS CloudFormation StackSets Per le strategie multi-sito attivo/attivo e standby a caldo, l'infrastruttura distribuita nella tua regione di ripristino ha le stesse risorse della regione principale. Per le strategie Pilot Light e Warm Standby l'infrastruttura distribuita richiederà azioni aggiuntive per essere pronta per la produzione. Utilizzando CloudFormation i parametri e la logica condizionale, puoi controllare se uno stack distribuito è attivo o in standby con un unico modello. Quando usi Elastic Disaster Recovery, il servizio replica e orchestra il ripristino delle configurazioni delle applicazioni e delle risorse di calcolo.

   Tutte le strategie di DR richiedono che venga eseguito il backup delle fonti di dati all'interno della regione di Regione AWS ripristino e quindi che tali backup vengano copiati nella regione di ripristino. AWS Backupoffre una visualizzazione centralizzata in cui è possibile configurare, pianificare e monitorare i backup per queste risorse. Per Pilot Light, Warm Standby e Multi-site active/active, è inoltre necessario replicare i dati dalla regione principale alle risorse di dati nella regione di ripristino, come le istanze RDSDB di Amazon Relational Database Service (Amazon) o le tabelle Amazon DynamoDB. Queste risorse di dati sono pertanto attive e pronte per servire le richieste nella regione di ripristino.

   Per saperne di più su come funzionano AWS i servizi in tutte le regioni, consulta questa serie di blog sulla creazione di un'applicazione multiregionale con servizi. AWS

4. Stabilisci e implementa le modalità con cui preparerai la tua regione al failover nel momento in cui sarà necessario (durante un'emergenza).

   Per la strategia attivo/attivo multisito, il failover significa evacuare una regione e usare le regioni attive rimanenti. In generale, tali regioni sono pronte per accettare il traffico. Per le strategie Pilot Light e Warm Standby, le azioni di ripristino dovranno distribuire le risorse mancanti, come le EC2 istanze nella Figura 20, oltre a qualsiasi altra risorsa mancante.

   Per tutte le strategie precedenti potresti dover promuovere istanze di database di sola lettura a istanze di lettura/scrittura principali.

   Per il backup e il ripristino, il ripristino dei dati dal backup crea risorse per tali dati come EBS volumi, istanze RDS DB e tabelle DynamoDB. Devi anche ripristinare l'infrastruttura e implementare il codice. È possibile utilizzare AWS Backup per ripristinare i dati nella regione di

ripristino. Per ulteriori dettagli, consulta REL09-BP01 Identifica ed esegui il backup di tutti i dati di cui è necessario eseguire il backup o riproduci i dati dalle fonti. La ricostruzione dell'infrastruttura include la creazione di risorse come EC2 istanze oltre all'Amazon Virtual Private Cloud VPC (Amazon), alle sottoreti e ai gruppi di sicurezza necessari. Puoi automatizzare gran parte del processo di ripristino. Per scoprire come farlo, consulta questo post del blog.

5. Stabilisci e implementa le modalità con cui reindirizzerai il traffico al failover nel momento in cui sarà necessario (durante un'emergenza).

   Questa operazione di failover può essere avviata automaticamente o manualmente. Il failover avviato automaticamente in base a controlli dell'integrità o allarmi deve essere usato con attenzione, poiché un failover non necessario (falso allarme) comporta dei costi in termini di non disponibilità e perdita dei dati. Pertanto si usa spesso il failover avviato manualmente. In questo caso, devi comunque automatizzare i passaggi del failover, in modo che l'avvio manuale si limiti al clic su un pulsante.

   Esistono diverse opzioni di gestione del traffico da considerare quando si utilizzano i servizi. AWS Tra le opzioni, vi è l'utilizzo di Amazon Route 53. Utilizzando Amazon Route 53, puoi associare più endpoint IP in uno o più Regioni AWS con un nome di dominio Route 53. Per implementare il failover avviato manualmente, puoi utilizzare Amazon Application Recovery Controller, che fornisce un piano dati ad alta disponibilità per API reindirizzare il traffico verso la regione di ripristino. Nella fase di implementazione del failover, usa le operazioni di piano dati ed evita quelle del piano di controllo (control-plane), come illustrato in REL11-BP04 Affidati al piano dati e non al piano di controllo durante il ripristino.

   Per ulteriori informazioni su questa e altre opzioni, consulta questa sezione del whitepaper sul ripristino di emergenza.

6. Progetta un piano per il failback del carico di lavoro.

   Si parla di failback quando un'operazione del carico di lavoro torna alla regione principale, dopo che un'emergenza è diminuita di intensità. Il provisioning di infrastruttura e codice alla regione principale in genere segue gli stessi passaggi usati inizialmente, affidandosi al modello Infrastructure as code e alle pipeline di implementazione del codice. La sfida del failback è il ripristino dei data store e la garanzia della loro coerenza con la regione di ripristino attiva.

   Nello stato di failover, i database nella regione di ripristino sono attivi e contengono i dati. up-to-date L'obiettivo è quindi quello di risincronizzare dalla regione di ripristino alla regione principale, assicurandosi che ciò avvenga. up-to-date

Alcuni AWS servizi lo faranno automaticamente. In caso di utilizzo delle tabelle globali Amazon DynamoDB, anche se la tabella nella regione principale era diventata non disponibile, quando torna di nuovo online, ripristina la propagazione di scritture in sospeso. Se utilizzi il Database globale Amazon Aurora e un failover pianificato gestito, viene mantenuta la topologia di replica esistente del database globale Aurora. Pertanto, l'istanza precedente in lettura/scrittura nella regione principale diventa una replica e riceve gli aggiornamenti dalla regione di ripristino.

Nei casi in cui questo non è automatico devi ristabilire il database nella regione principale come replica del database nella regione di ripristino. In molti casi questo comporterà l'eliminazione del database principale precedente e la creazione di nuove repliche.

Dopo un failover, se puoi proseguire l'esecuzione nella tua regione di ripristino, valuta la possibilità di farlo nella tua regione principale. Effettueresti comunque tutte le operazioni precedenti per trasformare la precedente regione principale in una regione di ripristino. Alcune organizzazioni eseguono una rotazione pianificata, scambiando periodicamente le regioni principale e di ripristino (ad esempio, ogni tre mesi).

Tutti i passaggi richiesti per failover e failback devono essere inseriti in un playbook disponibile a tutti i membri del team, sottoposto periodicamente a revisione.

Se usi Elastic Disaster Recovery, il servizio fornirà assistenza per l'orchestrazione e l'automazione del processo di failback. Per ulteriori informazioni, consulta Performing a failback.

Livello di impegno per il piano di implementazione: elevato

Risorse

Best practice correlate:

- the section called “REL09-BP01 Identifica ed esegui il backup di tutti i dati di cui è necessario eseguire il backup o riproduci i dati dalle fonti”
- the section called “REL11-BP04 Fai affidamento sul piano dati e non sul piano di controllo durante il ripristino”
- the section called “REL13-BP01 Definire gli obiettivi di ripristino per i tempi di inattività e la perdita di dati”

Documenti correlati:

- AWS Architecture Blog: serie sul ripristino di emergenza
- Disaster Recovery dei carichi di lavoro su AWS: Recovery in the Cloud (AWS Whitepaper)
- Opzioni di ripristino di emergenza nel cloud
- Build a serverless multi-region, active-active backend solution in an hour
- Multi-region serverless backend — reloaded
- RDS: Replica di una replica di lettura in diverse regioni
- Route 53: configurazione del failover DNS
- S3: replica tra regioni
- Che cos'è? AWS Backup
- Cos'è Amazon Application Recovery Controller?
- AWS Elastic Disaster Recovery
- HashiCorpTerraform: Guida introduttiva - AWS
- APNPartner: partner che possono aiutare con il disaster recovery
- Marketplace AWS: prodotti utilizzabili per il ripristino di emergenza

Video correlati:

- Disaster recovery dei carichi di lavoro su AWS
- AWS re:Invent 2018: modelli di architettura per applicazioni Active-Active in più regioni (09-R2) ARC2
- Inizia a usare AWS Elastic Disaster Recovery | Amazon Web Services

Esempi correlati:

- Well-Architected Lab: ripristino di emergenza, serie di workshop che illustrano le strategie di ripristino di emergenza

REL13-BP03 Testare l'implementazione del disaster recovery per convalidare l'implementazione

Testate regolarmente il failover sul sito di ripristino per verificare che funzioni correttamente RTO e RPO che sia rispettato.

Anti-pattern comuni:

- Non eseguire mai failover di prova in produzione.

Vantaggi dell'adozione di questa best practice: testare regolarmente il piano di ripristino di emergenza verifica che funzioni quando necessario e che il tuo team sappia come eseguire la strategia.

Livello di rischio associato se questa best practice non fosse adottata: elevato

Guida all'implementazione

Un modello da evitare è lo sviluppo di percorsi di ripristino eseguiti raramente. Ad esempio, è possibile che si disponga di un archivio dati secondario utilizzato per query di sola lettura. Quando scrivi in un archivio dati e quello principale ha un guasto, puoi eseguire il failover verso l'archivio dati secondario. Se non testi frequentemente questo failover, è possibile che i presupposti relativi alle funzionalità dell'archivio dati secondario non siano corretti. La capacità dell'archivio dati secondario, che potrebbe essere stata sufficiente durante l'ultimo test, potrebbe non essere più in grado di tollerare il carico in questo scenario. La nostra esperienza ha dimostrato che l'unico ripristino da errore che funziona è il percorso sottoposto a frequenti test. Per questo è preferibile avere un numero ridotto di percorsi di ripristino. Puoi stabilire dei modelli di ripristino e testarli regolarmente. Se disponi di un percorso di ripristino complesso o critico, devi comunque riprodurre regolarmente il guasto specifico in produzione per convincerti che il percorso di ripristino funzioni. Nell'esempio appena discusso, è necessario eseguire il failover regolarmente in standby, indipendentemente dalle necessità.

Passaggi dell'implementazione

1. Progetta i carichi di lavoro per il ripristino. Esegui regolarmente test dei tuoi percorsi di ripristino. Il calcolo orientato al ripristino identifica le caratteristiche nei sistemi che migliorano il ripristino: isolamento e ridondanza, ripristino a livello di sistema dello stato precedente rispetto alle modifiche, capacità di fornire diagnostica, ripristino automatico, progettazione modulare e possibilità di riavvio. Prova il percorso di ripristino per verificare di poter completare il ripristino nel tempo specificato e in base allo stato specificato. Usa i tuoi runbook durante questo ripristino per documentare i problemi e trovarne le soluzioni prima del test successivo.
2. Per i carichi di lavoro EC2 basati su Amazon, utilizzali AWS Elastic Disaster Recoveryper implementare e lanciare istanze drill per la tua strategia di disaster recovery. AWS Elastic Disaster Recovery offre la possibilità di eseguire esercitazioni in modo efficiente, il che aiuta a prepararsi per un evento di failover. Puoi anche avviare spesso le istanze usando Elastic Disaster Recovery per scopi di test ed esercitazione senza reindirizzare il traffico.

Risorse

Documenti correlati:

- APNPartner: partner che possono contribuire al disaster recovery
- AWS Architecture Blog: serie sul ripristino di emergenza
- Marketplace AWS: prodotti utilizzabili per il ripristino di emergenza
- AWS Elastic Disaster Recovery
- Disaster recovery dei carichi di lavoro su AWS: Recovery in the Cloud (AWS white paper)
- AWS Elastic Disaster Recovery Preparazione per il failover
- The Berkeley/Stanford recovery-oriented computing project
- Cos'è AWS Fault Injection Simulator?

Video correlati:

- AWS re:Invent 2018: modelli di architettura per applicazioni Active-Active in più regioni
- AWS re:Invent 2019: e soluzioni di disaster recovery con Backup-and-restore AWS

Esempi correlati:

- Well-Architected Labs: test di resilienza

REL13-BP04 Gestire la deriva della configurazione nel sito o nella regione di DR

Assicurati che l'infrastruttura, i dati e la configurazione soddisfino le esigenze del sito di ripristino di emergenza o della regione. Ad esempio, verifica che le quote di servizio AMIs e di servizio siano aggiornate.

AWS Config monitora e registra continuamente le configurazioni AWS delle risorse. È in grado di rilevare la deriva e richiamare AWS Systems Manager Automation per correggerla e generare allarmi. AWS CloudFormation può inoltre rilevare la deriva negli stack che hai distribuito.

Anti-pattern comuni:

- Non eseguire aggiornamenti nelle sedi di ripristino, quando esegui modifiche di configurazione o di infrastruttura nelle tue sedi principali.

- Ignorare le limitazioni potenziali (ad esempio le differenze di servizio) nelle sedi di disaster recovery e principali.

Vantaggi dell'adozione di questa best practice: assicurarsi che l'ambiente di ripristino di emergenza sia coerente con quello esistente garantisce il ripristino completo.

Livello di rischio associato se questa best practice non fosse adottata: medio

Guida all'implementazione

- Assicurati che le tue pipeline di distribuzione riforniscano sia i siti principali che di backup. Le pipeline per l'implementazione di applicazioni in produzione devono essere distribuite in tutte le posizioni della strategia di ripristino di emergenza specificate, inclusi gli ambienti di sviluppo e test.
- Permesso di tracciare AWS Config i potenziali punti di deriva. Utilizza AWS Config le regole per creare sistemi che applichino le tue strategie di disaster recovery e generino avvisi quando rilevano deviazioni.
  - Risanamento delle risorse non conformi mediante AWSRegole di AWS Config
  - AWS Systems Manager Automation
- AWS CloudFormation Utilizzalo per implementare la tua infrastruttura. AWS CloudFormation è in grado di rilevare differenze tra ciò che viene specificato CloudFormation dai modelli e ciò che viene effettivamente distribuito.
  - AWS CloudFormation: rileva la deriva su un intero stack CloudFormation

Risorse

Documenti correlati:

- APNPartner: partner che possono contribuire al disaster recovery
- AWS Architecture Blog: serie sul ripristino di emergenza
- AWS CloudFormation: Rileva la deriva su un intero stack CloudFormation
- Marketplace AWS: prodotti utilizzabili per il ripristino di emergenza
- AWS Systems Manager Automation
- Disaster recovery dei carichi di lavoro su AWS: Recovery in the Cloud (AWS white paper)
- In che modo è possibile implementare una soluzione di gestione della configurazione dell'infrastruttura in AWS?

- Risanamento delle risorse non conformi mediante AWSRegole di AWS Config

Video correlati:

- AWS re:Invent 2018: modelli di architettura per applicazioni Active-Active in più regioni (09-R2) ARC2

REL13-BP05 Ripristino automatico

Utilizza strumenti AWS di terze parti per automatizzare il ripristino del sistema e indirizzare il traffico verso il sito o la regione di DR.

In base ai controlli di integrità configurati, AWS i servizi, come Elastic Load Balancing e AWS Auto Scaling, possono distribuire il carico verso zone di disponibilità integre, mentre i servizi, come Amazon Route 53 e AWS Global Accelerator, possono indirizzare il carico verso l'integrità. Regioni AWS Amazon Application Recovery Controller ti aiuta a gestire e coordinare il failover utilizzando le funzionalità di controllo della disponibilità e del routing. Queste funzionalità monitorano continuamente la capacità dell'applicazione di ripristinarsi in caso di guasti, in modo da poter controllare il ripristino delle applicazioni su più Regioni AWS zone di disponibilità e in locale.

Per i carichi di lavoro su data center fisici o virtuali esistenti o cloud privati, AWS Elastic Disaster Recovery consente alle organizzazioni di configurare una strategia di ripristino di emergenza automatizzata in AWS. Elastic Disaster Recovery supporta anche il ripristino di emergenza tra regioni e zone di disponibilità in AWS.

Anti-pattern comuni:

- L'implementazione di failover e failback automatici identici può causare flapping quando si verifica un errore.

Vantaggi dell'adozione di questa best practice: il ripristino automatico riduce i tempi di ripristino eliminando la possibilità di errori manuali.

Livello di rischio associato se questa best practice non fosse adottata: medio

Guida all'implementazione

- Automatizzazione dei percorsi di ripristino. Per tempi di ripristino brevi, segui il tuo piano di ripristino di emergenza così da riportare rapidamente online i sistemi IT in caso di interruzione.

- Utilizza Elastic Disaster Recovery per failover e failback automatizzati Elastic Disaster Recovery replica continuamente le tue macchine (inclusi sistema operativo, configurazione dello stato del sistema, database, applicazioni e file) in un'area di staging a basso costo nella tua regione di destinazione e preferita. Account AWS In caso di emergenza, una volta scelto il ripristino mediante Elastic Disaster Recovery, la soluzione automatizza la conversione dei server replicati in carichi di lavoro completamente allocati nella regione di ripristino attiva in AWS.
  - Using Elastic Disaster Recovery for Failover and Failback
  - AWS Elastic Disaster Recovery resources

Risorse

Documenti correlati:

- APNPartner: partner che possono aiutarti con il disaster recovery
- AWS Architecture Blog: serie sul ripristino di emergenza
- Marketplace AWS: prodotti utilizzabili per il ripristino di emergenza
- AWS Systems Manager Automation
- AWS Elastic Disaster Recovery
- Disaster recovery dei carichi di lavoro su AWS: Recovery in the Cloud (AWS white paper)

Video correlati:

- AWS re:Invent 2018: modelli di architettura per applicazioni Active-Active in più regioni (09-R2) ARC2

Efficienza delle prestazioni

Il pilastro dell'efficienza delle prestazioni include la capacità di utilizzare in modo efficiente le risorse nel cloud per soddisfare i requisiti in termini di prestazione e di mantenere tale efficienza a fronte al cambiamento della domanda e all'evoluzione delle tecnologie. Puoi trovare linee guida prescrittive sull'implementazione nel whitepaper sul pilastro dell'efficienza delle prestazioni.

Aree delle best practice

- Scelta dell'architettura
- Calcolo e hardware

- Gestione dei dati
- Reti e distribuzione di contenuti
- Processo e cultura

Scelta dell'architettura

Questions

- PERF1. In che modo selezioni le risorse e l'architettura cloud appropriate per il tuo carico di lavoro?

PERF1. In che modo selezioni le risorse e l'architettura cloud appropriate per il tuo carico di lavoro?

La soluzione ottimale per un determinato carico di lavoro può variare e le soluzioni spesso combinano molteplici approcci. I carichi di lavoro Well-Architected utilizzano soluzioni multiple e forniscono funzionalità diverse per migliorare le prestazioni.

Best practice

- PERF01-BP01 Scopri e comprendi i servizi e le funzionalità cloud disponibili
- PERF01-BP02 Utilizza la guida del tuo provider di servizi cloud o di un partner appropriato per conoscere i modelli di architettura e le migliori pratiche
- PERF01-BP03 Fattore di costo nelle decisioni architettoniche
- PERF01-BP04 Valuta l'impatto dei compromessi sui clienti e sull'efficienza dell'architettura
- PERF01-BP05 Usa politiche e architetture di riferimento
- PERF01-BP06 Usa il benchmarking per guidare le decisioni sull'architettura
- PERF01-BP07 Usa un approccio basato sui dati per le scelte architettoniche

PERF01-BP01 Scopri e comprendi i servizi e le funzionalità cloud disponibili

Informati continuamente e identifica i servizi e le configurazioni disponibili che ti aiutano a prendere le decisioni giuste sull'architettura e a migliorare l'efficienza delle prestazioni dei carichi di lavoro.

Anti-pattern comuni:

- Utilizzi il cloud come data center in co-location.

- Non stai modernizzando la tua applicazione con la migrazione al cloud.
- Stai solo usando un tipo di archiviazione per tutte le cose che devono essere conservate in modo persistente.
- Se necessario, utilizzi tipi di istanze strettamente correlate ai tuoi standard attuali, ma più grandi.
- Distribuisci e gestisci le tecnologie disponibili come servizi gestiti.

Vantaggi dell'adozione di questa best practice: prendendo in considerazione nuovi servizi e configurazioni, puoi migliorare notevolmente le prestazioni, ridurre i costi e ottimizzare le attività necessarie per mantenere il carico di lavoro. Può anche aiutarti ad accelerare l'adozione di prodotti abilitati al cloud time-to-value.

Livello di rischio associato se questa best practice non fosse adottata: elevato

Guida all'implementazione

AWS rilascia continuamente nuovi servizi e funzionalità in grado di migliorare le prestazioni e ridurre il costo dei carichi di lavoro cloud. Utilizzare up-to-date questi nuovi servizi e funzionalità è fondamentale per mantenere l'efficacia delle prestazioni nel cloud. La modernizzazione dell'architettura dei carichi di lavoro consente inoltre di accelerare la produttività, promuovere l'innovazione e sbloccare ulteriori opportunità di crescita.

Passaggi dell'implementazione

- Esegui l'inventario del software e dell'architettura del carico di lavoro per i servizi correlati. Determina su quale categoria di prodotti ottenere ulteriori informazioni.
- Esplora AWS le offerte per identificare e conoscere i servizi e le opzioni di configurazione pertinenti che possono aiutarti a migliorare le prestazioni e ridurre i costi e la complessità operativa.
  - Amazon Web Services Cloud
  - AWS Accademia
  - Cosa c'è di nuovo con AWS?
  - AWS Blog
  - AWS Skill Builder
  - AWS Eventi e webinar
  - AWS Training e certificazioni
  - AWS Canale Youtube

  - AWS Workshop
  - Community AWS
- Usa Amazon Q per ricevere informazioni e consigli pertinenti sui servizi.
- Usa gli ambienti sandbox non di produzione per comprendere e sperimentare nuovi servizi senza incorrere in costi aggiuntivi.
- Scopri servizi e funzionalità cloud sempre nuovi.

Risorse

Documenti correlati:

- Overview of Amazon Web Services
- EC2Funzionalità di Amazon
- Impara step-by-step con un piano formativo per i AWS partner
- AWS Formazione e certificazione
- Il mio percorso di apprendimento per diventare un architetto di AWS soluzioni
- AWS Centro di architettura
- AWS Partner Network
- AWS Libreria di soluzioni
- AWS Centro di conoscenza
- Crea applicazioni moderne su AWS

Video correlati:

- AWS re:Invent 2023 - Cosa c'è di nuovo con Amazon EC2
- AWS re:Invent 2022 - Riduci i costi operativi e di infrastruttura con Amazon ECS
- AWS re:Invent 2023 - Costruisci con l'efficienza, l'agilità e l'innovazione del cloud con AWS
- AWS re:Invent 2022 - Implementa modelli ML per l'inferenza ad alte prestazioni e basso costo
- This is my Architecture

Esempi correlati:

- AWS Esempi

- AWS SDKEsempi

PERF01-BP02 Utilizza la guida del tuo provider di servizi cloud o di un partner appropriato per conoscere i modelli di architettura e le migliori pratiche

Usa le risorse aziendali del cloud come documentazione, solutions architect, servizi professionali o partner appropriati per guidare le tue decisioni sull'architettura. Queste risorse ti aiutano a rivedere e migliorare l'architettura per ottenere prestazioni ottimali.

Anti-pattern comuni:

- Lo utilizzi AWS come provider di servizi cloud comune.
- Utilizzi AWS i servizi in un modo per cui non sono stati progettati.
- Le indicazioni vengono seguite senza considerare il contesto aziendale.

Vantaggi dell'adozione di questa best practice: avvalersi della guida di un provider di servizi cloud o di un partner appropriato può aiutarti a fare le scelte giuste per l'architettura del tuo carico di lavoro e darti fiducia nelle tue decisioni.

Livello di rischio associato se questa best practice non fosse adottata: medio

Guida all'implementazione

AWS offre un'ampia gamma di indicazioni, documentazione e risorse che possono aiutarti a creare e gestire carichi di lavoro cloud efficienti. AWS la documentazione fornisce esempi di codice, tutorial e spiegazioni dettagliate dei servizi. Oltre alla documentazione, AWS fornisce programmi di formazione e certificazione, architetti di soluzioni e servizi professionali che possono aiutare i clienti a esplorare diversi aspetti dei servizi cloud e implementare un'architettura cloud efficiente su. AWS

Sfrutta queste risorse per ottenere approfondimenti sulle informazioni e sulle best practice preziose per risparmiare tempo e ottenere risultati migliori nel Cloud AWS.

Passaggi dell'implementazione

- AWS Consulta la documentazione e le linee guida e segui le migliori pratiche. Queste risorse possono aiutarti a scegliere e configurare i servizi in modo efficace e a ottenere prestazioni migliori.
  - AWS documentazione (come guide per l'utente e white paper)
  - AWS Blog
  - AWS Training e certificazioni

  - AWS Canale Youtube
- Partecipa agli eventi dei AWS partner (come AWS Global Summits, AWS re:Invent, gruppi di utenti e workshop) per imparare dagli AWS esperti le migliori pratiche per l'utilizzo dei servizi. AWS
  - Impara step-by-step con un piano formativo per i partner AWS
  - AWS Eventi e webinar
  - AWS Workshop
  - AWS Comunità
- Rivolgiti a AWS noi per ricevere assistenza quando hai bisogno di ulteriori indicazioni o informazioni sul prodotto. AWS Solutions Architects e AWS Professional Services forniscono indicazioni per l'implementazione della soluzione. AWS I partner forniscono AWS competenze per aiutarvi a sbloccare l'agilità e l'innovazione per la vostra azienda.
- Usa AWS Support se hai bisogno di supporto tecnico per utilizzare un servizio in modo efficace. I nostri piani di supporto sono progettati per offrirti il giusto mix di strumenti e l'accesso alle competenze in modo che tu possa avere successo ottimizzando al AWS contempo le prestazioni, gestendo i rischi e mantenendo i costi sotto controllo.

Risorse

Documenti correlati:

- AWS Architecture Center
- AWS Partner Network
- Biblioteca di soluzioni di AWS
- Centro conoscenze di AWS
- Supporto AWS Enterprise

Video correlati:

- This is my Architecture
- AWS re:Invent 2023 - Modelli avanzati basati sugli eventi con Amazon EventBridge
- AWS re:Invent 2023 - Implementazione di modelli di progettazione distribuiti su AWS
- AWS re:Invent 2023 - Architettura dell'applicazione come codice

Esempi correlati:

- Esempi AWS
- AWS SDKEsempi
- AWS Architettura di riferimento di Analytics

PERF01-BP03 Fattore di costo nelle decisioni architettoniche

Tieni conto dei costi nelle decisioni sull'architettura per migliorare l'utilizzo delle risorse e l'efficienza delle prestazioni del tuo carico di lavoro cloud. Quando si è consapevoli delle implicazioni dei costi del carico di lavoro cloud, è più probabile che si utilizzino risorse efficienti e si riducano le procedure inutili.

Anti-pattern comuni:

- Utilizzi una sola famiglia di istanze.
- Ometti di valutare le soluzioni con licenza rispetto alle soluzioni open-source.
- Non definisci le policy del ciclo di vita dell'archiviazione.
- Non recensisci i nuovi servizi e funzionalità di. Cloud AWS
- Utilizzi solo lo storage a blocchi.

Vantaggi dell'adozione di questa best practice: la contabilizzazione dei costi nel processo decisionale consente di utilizzare risorse più efficienti ed esplorare altri investimenti.

Livello di rischio associato se questa best practice non fosse adottata: medio

Guida all'implementazione

L'ottimizzazione dei carichi di lavoro in base ai costi può migliorare l'utilizzo delle risorse ed evitare sprechi nel carico di lavoro cloud. Tenere conto dei costi nelle decisioni sull'architettura di solito include il corretto dimensionamento dei componenti del carico di lavoro e l'abilitazione dell'elasticità, comportando una migliore efficienza delle prestazioni del carico di lavoro cloud.

Passaggi dell'implementazione

- Stabilisci gli obiettivi di costo, come i limiti del budget, per il tuo carico di lavoro cloud.
- Identifica i componenti chiave, come istanze e archiviazione, che determinano il costo del carico di lavoro. Puoi usare AWS Pricing Calculator e AWS Cost Explorer per identificare i principali fattori di costo del carico di lavoro.

- Esamina i modelli di prezzo nel cloud, ad esempio istanze on-demand, riservate, Savings Plans e istanze spot.
- Segui le best practice per l'ottimizzazione dei costi di Well-Architected per ottimizzare questi componenti principali in termini di costi.
- Monitora e analizza continuamente i costi per identificare le opportunità di ottimizzazione dei costi nel tuo carico di lavoro.
  - Usa Budget AWS per ricevere gli avvisi per i costi inaccettabili.
  - Usa AWS Compute Optimizer o AWS Trusted Advisor per ottenere suggerimenti sull'ottimizzazione dei costi.
  - Usa AWS Cost Anomaly Detection per rilevare in modo automatico le anomalie dei costi e analizzare la causa principale.

Risorse

Documenti correlati:

- Che cos'è AWS Billing and Cost Management?
- Ottimizzazione dei costi con AWS
- Scelta di una strategia di gestione dei AWS costi
- Una guida per principianti alla gestione AWS dei costi
- A Detailed Overview of the Cost Intelligence Dashboard
- AWS Architecture Center
- Biblioteca di soluzioni di AWS
- Centro conoscenze di AWS

Video correlati:

- This is my Architecture
- AWS re:Invent 2023 - Cosa c'è di nuovo con l'ottimizzazione dei costi AWS
- AWS re:Invent 2023 - Ottimizza costi e prestazioni e monitora i progressi verso la mitigazione
- AWS re:Invent 2023 - best practice per l'ottimizzazione dei costi di storage AWS
- AWS re:Invent 2023 - Ottimizza i costi nei tuoi ambienti con più account

Esempi correlati:

- AWS Compute Optimizer Codice demo
- Cost Optimization Workshop
- Cloud Financial Management Technical Implementation Playbooks
- Startup optimization: Tuning application performance for maximum efficiency
- Serverless Optimization Workshop (Performance and Cost)
- Scaling cost effective architectures

PERF01-BP04 Valuta l'impatto dei compromessi sui clienti e sull'efficienza dell'architettura

Quando valuti i miglioramenti correlati alle prestazioni, determina quali scelte hanno impatto sui clienti e sull'efficienza del carico di lavoro. Ad esempio, se l'utilizzo di un datastore chiave-valore aumenta le prestazioni del sistema, è importante valutare in che modo la consistenza finale intrinseca di questo cambiamento avrà un impatto sui clienti.

Anti-pattern comuni:

- Ritieni che tutti i vantaggi prestazionali debbano essere implementati, anche se ci sono compromessi per l'implementazione.
- Valuti di apportare modifiche ai carichi di lavoro solo quando un problema prestazionale ha raggiunto un punto critico.

Vantaggi dell'adozione di questa best practice: quando si valutano potenziali miglioramenti relativi alle prestazioni, è necessario decidere se i compromessi per le modifiche sono accettabili con i requisiti del carico di lavoro. In alcuni casi, potrebbe essere necessario implementare controlli aggiuntivi per compensare i compromessi.

Livello di rischio associato se questa best practice non fosse adottata: elevato

Guida all'implementazione

Identifica le aree critiche della tua architettura in termini di prestazioni e impatto sui clienti. Stabilisci in che modo puoi apportare miglioramenti e quali compromessi comportano, oltre al loro impatto sul sistema e sull'esperienza degli utenti. L'implementazione di cache di dati, ad esempio, può contribuire a migliorare notevolmente le prestazioni ma richiede una strategia ben definita sulle modalità e sui

tempi di aggiornamento o di invalidamento dei dati che vi sono contenuti, per evitare che il sistema si comporti in modo non corretto.

Passaggi dell'implementazione

- Comprendi i requisiti del tuo carico di SLAs lavoro e.
- Definisci chiaramente i fattori di valutazione. I fattori possono riguardare il costo, l'affidabilità, la sicurezza e le prestazioni del carico di lavoro.
- Seleziona l'architettura e i servizi in grado di soddisfare le tue esigenze.
- Conduci sperimentazioni e prove di fattibilità (POCs) per valutare i fattori di compromesso e l'impatto sui clienti e sull'efficienza dell'architettura. Di solito, i carichi di lavoro altamente disponibili, performanti e sicuri consumano più risorse cloud offrendo al contempo una esperienza cliente migliore. Comprendi i compromessi in termini di complessità, prestazioni e costi del tuo carico di lavoro. In genere, dare la priorità a due fattori va a scapito del terzo.

Risorse

Documenti correlati:

- Amazon Builders' Library
- Amazon QuickSight KPIs
- Amazon CloudWatch RUM
- Documentazione di X-Ray
- Understand resiliency patterns and trade-offs to architect efficiently in the cloud

Video correlati:

- Ottimizza le applicazioni tramite Amazon CloudWatch RUM
- AWS re:Invent 2023 - Capacità, disponibilità, efficienza dei costi: scegline tre
- AWS re:Invent 2023 - Modelli di integrazione avanzati e compromessi per sistemi liberamente accoppiati

Esempi correlati:

- Misura il tempo di caricamento della pagina con Amazon CloudWatch Synthetics

- Client CloudWatch RUM Web Amazon

PERF01-BP05 Usa politiche e architetture di riferimento

Utilizza le policy interne e le architetture di riferimento esistenti per la selezione dei servizi e delle configurazioni per una maggiore efficienza nella progettazione e nell'implementazione del carico di lavoro.

Anti-pattern comuni:

- Usi una vasta gamma di tecnologie che possono influire sul sovraccarico di gestione della tua azienda.

Vantaggi dell'adozione di questa best practice: la definizione di una policy per la scelta dell'architettura, della tecnologia e del fornitore consente di prendere decisioni rapidamente.

Livello di rischio associato se questa best practice non fosse adottata: medio

Guida all'implementazione

Avere policy interne nella selezione delle risorse e dell'architettura fornisce standard e linee guida da seguire quando si effettuano scelte architettoniche. Queste linee guida semplificano il processo decisionale nella scelta del servizio cloud giusto e possono contribuire a migliorare l'efficienza delle prestazioni. Implementi il carico di lavoro utilizzando policy o architetture di riferimento. Integra i servizi nell'implementazione cloud, quindi utilizza i test delle prestazioni per verificare che i requisiti prestazionali siano sempre rispettati.

Passaggi dell'implementazione

- Comprendi chiaramente i requisiti del tuo carico di lavoro cloud.
- Rivedi le policy interne ed esterne per identificare quelle più pertinenti.
- Utilizza le architetture di riferimento appropriate fornite dalle best practice AWS o di settore.
- Crea un contesto composto da policy, standard, architetture di riferimento e linee guida prescrittive per situazioni comuni. In questo modo i tuoi team possono muoversi più velocemente. Personalizza le risorse per il tuo settore verticale, se applicabile.
- Convalida queste policy e architetture di riferimento per il tuo carico di lavoro in ambienti sandbox.
- Resta up-to-date conforme agli standard e agli AWS aggiornamenti del settore per assicurarti che le tue policy e le architetture di riferimento contribuiscano a ottimizzare il carico di lavoro sul cloud.

Risorse

Documenti correlati:

- AWS Architecture Center
- AWS Partner Network
- Biblioteca di soluzioni di AWS
- Centro conoscenze di AWS
- AWS Blog di architettura

Video correlati:

- This is my Architecture
- AWS re:Invent 2022 - Accelera il valore della tua azienda con SAP un'architettura di riferimento AWS

Esempi correlati:

- Esempi AWS
- AWS SDKEsempi

PERF01-BP06 Usa il benchmarking per guidare le decisioni sull'architettura

Esegui il benchmark delle prestazioni di un carico di lavoro esistente per comprendere le prestazioni sul cloud e guidare le decisioni sull'architettura basate sui dati.

Anti-pattern comuni:

- Fai affidamento su valori di riferimento comuni che non sono indicativi delle caratteristiche del carico di lavoro.
- L'unico punto di riferimento è dato dal feedback e dalle percezioni dei clienti.

Vantaggi dell'adozione di questa best practice: misurazione dei miglioramenti in termini di prestazioni grazie al benchmarking dell'implementazione attuale.

Livello di rischio associato se questa best practice non fosse adottata: medio

Guida all'implementazione

Utilizza test sintetici di benchmarking per valutare le prestazioni dei componenti durante il carico di lavoro. Di solito, i benchmark sono più rapidi da configurare rispetto ai test di carico e vengono utilizzati per valutare la tecnologia di un componente specifico. Il benchmarking viene spesso utilizzato all'inizio di un nuovo progetto, quando non è ancora disponibile una soluzione completa da sottoporre a test di carico.

Puoi creare test di benchmark personalizzati o utilizzare un test standard del settore, come TPC-DS, per confrontare i tuoi carichi di lavoro. I benchmark di settore sono utili quando devi confrontare ambienti diversi. Quelli personalizzati, invece, sono indicati per analizzare tipi specifici di operazioni che prevedi di eseguire nell'architettura.

In fase di benchmarking, è importante effettuare delle operazioni preliminari sull'ambiente di test al fine di garantire la validità dei risultati. Dovrai eseguire lo stesso benchmark più volte, per verificare di avere acquisito ogni eventuale variazione nel corso del tempo.

Dal momento che, di solito, l'esecuzione dei benchmark è più rapida di quella dei test di carico, il benchmarking può essere utilizzato sin dalle prime fasi della pipeline di implementazione, così da fornire al team feedback più rapidi sulle deviazioni delle prestazioni. Quando valuti un cambiamento significativo in un componente o servizio, i benchmark possono essere un modo rapido per verificare se l'impegno necessario per apportare la modifica sia giustificato. L'utilizzo del benchmarking in combinazione con i test di carico è importante perché questi ultimi forniscono indicazioni sulle prestazioni del carico di lavoro in fase di produzione.

Passaggi dell'implementazione

- Pianifica e definisci:
  - Definisci gli obiettivi, la linea di base, gli scenari di test, le metriche (come CPU utilizzo, latenza o velocità effettiva) e il benchmark. KPIs
  - Concentrati sui requisiti degli utenti in termini di esperienza utente e su fattori come i tempi di risposta e l'accessibilità.
  - Individua uno strumento di benchmark adatto al tuo carico di lavoro. Puoi utilizzare AWS servizi come Amazon CloudWatch o uno strumento di terze parti compatibile con il tuo carico di lavoro.
- Configura ed esegui l'instrumentazione:
  - Imposta il tuo ambiente e configura le risorse.
  - Implementa il monitoraggio e la creazione di log per acquisire i risultati dei test.

- Esegui i test di benchmark e monitora:
  - Esegui i test di benchmark e monitora i parametri durante il test.
- Analizza e documenta:
  - Documenta il processo di benchmark e gli esiti.
  - Analizza i risultati per identificare i colli di bottiglia, le tendenze e le aree di miglioramento.
  - Usa i risultati dei test per prendere decisioni sull'architettura e modificare il carico di lavoro. Questa operazione può includere la modifica dei servizi o l'adozione di nuove funzionalità.
- Ottimizza e ripeti:
  - Modifica le configurazioni e le allocazioni delle risorse in base ai tuoi benchmark.
  - Ripeti il test del carico di lavoro dopo i cambiamenti per convalidare i miglioramenti.
  - Documenta le informazioni e ripeti il processo per identificare altre aree di miglioramento.

Risorse

Documenti correlati:

- AWS Centro di architettura
- AWS Partner Network
- AWS Libreria di soluzioni
- AWS Centro di conoscenza
- Amazon CloudWatch RUM
- Amazon CloudWatch Synthetics
- Genomics workflows, Part 5: automated benchmarking
- Confronta e ottimizza la distribuzione degli endpoint in Amazon SageMaker JumpStart

Video correlati:

- AWS re:Invent 2023 - Analisi comparativa delle partenze a freddo AWS Lambda
- Benchmarking stateful services in the cloud
- This is my Architecture
- Ottimizza le applicazioni tramite Amazon CloudWatch RUM
- Demo di Amazon CloudWatch Synthetics

Esempi correlati:

- AWS Esempi
- AWS SDKEsempi
- Test del carico distribuito
- Misura il tempo di caricamento della pagina con Amazon CloudWatch Synthetics
- Client CloudWatch RUM Web Amazon

PERF01-BP07 Usa un approccio basato sui dati per le scelte architettoniche

Definisci un approccio chiaro e basato sui dati per le scelte dell'architettura e verificare che vengano utilizzati i servizi e le configurazioni cloud corretti per soddisfare le tue esigenze aziendali specifiche.

Anti-pattern comuni:

- Ritieni che l'architettura corrente diventi statica e non venga aggiornata nel corso del tempo.
- Le tue scelte dell'architettura si basano su ipotesi e supposizioni.
- Introduci modifiche all'architettura nel tempo senza giustificazioni.

Vantaggi dell'adozione di questa best practice: con un approccio ben definito per le scelte dell'architettura, utilizzi i dati per influenzare la progettazione del carico di lavoro e prendere decisioni informate nel tempo.

Livello di rischio associato se questa best practice non fosse adottata: medio

Guida all'implementazione

Affidati all'esperienza e alle competenze interne in materia di cloud o utilizza risorse esterne, come casi d'uso pubblicati o whitepaper, per scegliere risorse e servizi per la tua architettura. È necessario definire con cura un processo che incoraggi la sperimentazione e il benchmarking con i servizi che possono essere utilizzati nel carico di lavoro.

I backlog dei carichi di lavoro critici devono consistere non solo in storie che offrono funzionalità rilevanti per l'azienda e gli utenti, ma anche in storie tecniche che definiscono la presentazione dell'architettura per il carico di lavoro. Questa presentazione include i nuovi progressi tecnologici e i nuovi servizi e li adotta sulla base di dati e giustificazioni adeguate. Verifica che l'architettura sia a prova di futuro e non diventi obsoleta.

Passaggi dell'implementazione

- Interagisci con le principali parti interessate per definire i requisiti del carico di lavoro, comprese le prestazioni, la disponibilità e le considerazioni sui costi. Includi fattori quali il numero di utenti e il modello di utilizzo del tuo carico di lavoro.
- Crea una presentazione dell'architettura o un backlog tecnologico a cui venga assegnata la priorità insieme al backlog funzionale.
- Valuta e identifica i diversi servizi cloud (per ulteriori dettagli, consulta PERF01-BP01 Scopri e comprendi i servizi e le funzionalità cloud disponibili).
- Esplora i diversi modelli di architettura, come microservizi o serverless, che soddisfano i tuoi requisiti di prestazioni (per maggiori dettagli, consulta PERF01-BP02 Utilizza la guida del tuo provider di servizi cloud o di un partner appropriato per conoscere i modelli di architettura e le migliori pratiche).

- Consulta altri team, diagrammi di architettura e risorse, come AWS Solution Architects, AWS Architecture Center e AWS Partner Network, per aiutarti a scegliere l'architettura giusta per il tuo carico di lavoro.

- Definisci i parametri, come il throughput e il tempo di risposta, che possono aiutarti a valutare le prestazioni del tuo carico di lavoro.
- Sperimenta e utilizza i parametri definiti per convalidare le prestazioni dell'architettura selezionata.
- Monitora continuamente e apporta le modifiche necessarie per mantenere ottimali le prestazioni della tua architettura.
- Documenta l'architettura e le decisioni selezionate come riferimento per aggiornamenti e apprendimenti futuri.
- Rivedi e aggiorna continuamente l'approccio di selezione dell'architettura in base agli apprendimenti, alle nuove tecnologie e ai parametri che indicano un problema o un cambiamento necessario nell'approccio attuale.

Risorse

Documenti correlati:

- Biblioteca di soluzioni di AWS
- Centro conoscenze di AWS

- Modelli architettonici su cui creare applicazioni basate End-to-End sui dati AWS

Video correlati:

- This is my Architecture
- AWS re:Invent 2021 - Impresa basata sui dati: passare dalla visione al valore
- AWS re:Invent 2022 - Fornire architetture sostenibili e ad alte prestazioni
- AWS re:Invent 2023 - Ottimizza costi e prestazioni e monitora i progressi verso la mitigazione
- AWS re:Invent 2022 - AWS ottimizzazione: misure attuabili per risultati immediati

Esempi correlati:

- Esempi AWS
- AWS SDKEsempi

Calcolo e hardware

Questions

- PERF2. In che modo selezioni e utilizzi le risorse di elaborazione nel tuo carico di lavoro?

PERF2. In che modo selezioni e utilizzi le risorse di elaborazione nel tuo carico di lavoro?

La soluzione ottimale in termini di calcolo per un determinato carico di lavoro potrebbe variare in base alla progettazione dell'applicazione, ai modelli di utilizzo e alle impostazioni di configurazione. Le architetture possono utilizzare diverse soluzioni di calcolo per vari componenti e impiegare funzionalità diverse per migliorare le prestazioni. Selezionare la soluzione di calcolo sbagliata per un'architettura può ridurre l'efficienza delle prestazioni.

Best practice

- PERF02-BP01 Seleziona le migliori opzioni di elaborazione per il tuo carico di lavoro
- PERF02-BP02 Comprendi la configurazione e le funzionalità di elaborazione disponibili
- PERF02-BP03 Raccogli metriche relative al calcolo
- PERF02-BP04 Configurazione e dimensionamento corretto delle risorse di elaborazione

- PERF02-BP05 Scala le tue risorse di elaborazione in modo dinamico
- PERF02-BP06 Usa acceleratori di calcolo ottimizzati basati su hardware

PERF02-BP01 Seleziona le migliori opzioni di elaborazione per il tuo carico di lavoro

La selezione dell'opzione di elaborazione più appropriata per il carico di lavoro consente di migliorare le prestazioni, ridurre i costi non necessari dell'infrastruttura e diminuire le attività operative richieste per mantenere il carico di lavoro.

Anti-pattern comuni:

- Si utilizza la stessa opzione di elaborazione utilizzata on-premises.
- Non si conoscono le opzioni, le funzionalità e le soluzioni di cloud computing e come queste migliorino le prestazioni di elaborazione.
- Si effettua il provisioning eccessivo dell'opzione di elaborazione per soddisfare i requisiti di dimensionamento o prestazioni, quando il passaggio a una nuova opzione di elaborazione soddisferebbe le caratteristiche del carico di lavoro in modo più preciso.

Vantaggi dell'adozione di questa best practice: identificando i requisiti di elaborazione e valutando le opzioni disponibili è possibile rendere il carico di lavoro più efficiente in termini di risorse.

Livello di rischio associato se questa best practice non fosse adottata: elevato

Guida all'implementazione

Per ottimizzare i carichi di lavoro cloud per l'efficienza delle prestazioni, è importante selezionare le opzioni di elaborazione più appropriate per il caso d'uso e i requisiti di prestazioni. AWS offre una varietà di opzioni di elaborazione che soddisfano diversi carichi di lavoro nel cloud. Ad esempio, puoi utilizzare Amazon EC2 per avviare e gestire server virtuali, AWS Lambdaeseguire codice senza dover fornire o gestire server, Amazon ECS o Amazon EKS per eseguire e gestire contenitori o AWS Batchelaborare grandi volumi di dati in parallelo. In base alle tue esigenze di dimensionamento ed elaborazione, scegli e configura la soluzione di elaborazione ottimale per la tua situazione. Puoi anche prendere in considerazione l'utilizzo di più tipi di soluzioni di elaborazione in un unico carico di lavoro in quanto ognuna ha i suoi vantaggi e svantaggi.

I passaggi seguenti ti guidano nella selezione delle opzioni di elaborazione giuste per soddisfare le caratteristiche del carico di lavoro e i requisiti prestazionali.

Passaggi dell'implementazione

- Comprendi i requisiti di elaborazione del tuo carico di lavoro. I requisiti essenziali da considerare includono le esigenze di elaborazione, gli schemi di traffico, gli schemi di accesso ai dati, le esigenze di dimensionamento e i requisiti di latenza.
- Scopri i vari servizi di elaborazione AWS per il tuo carico di lavoro. Per ulteriori informazioni, consulta PERF01-BP01 Scopri e comprendi i servizi e le funzionalità cloud disponibili. Ecco alcune importanti opzioni di elaborazione AWS , le caratteristiche e i casi d'uso più comuni:

AWS servizio	Caratteristiche chiave	Casi di utilizzo comune
Amazon Elastic Compute Cloud (AmazonEC2)	Dispone di un'opzione dedicata per hardware, requisiti di licenza, ampia selezione di diverse famiglie di istanze, tipi di processori e acceleratori di elaborazione	Migrazioni con rehosting (lift and shift), applicazione monolitica, ambienti ibridi, applicazioni aziendali
Amazon Elastic Container Service (AmazonECS), Amazon Elastic Kubernetes Service (Amazon) EKS	Implementazione semplice, ambienti coerenti, scalabile	Microservizi, ambienti ibridi
AWS Lambda	Servizio di elaborazione serverless che esegue il codice in risposta agli eventi e gestisce automaticamente le risorse di elaborazione sottostanti.	Microservizi, applicazioni basate su eventi
AWS Batch	Esegue il provisioning e la scalabilità in modo efficient e e dinamico di Amazon Elastic Container Service (AmazonECS), Amazon Elastic Kubernetes Service EKS (Amazon AWS Fargate)	HPC, addestra modelli ML

AWS servizio	Caratteristiche chiave	Casi di utilizzo comune
	e risorse di calcolo, con la possibilità di utilizzare istanze On-Demand o Spot in base alle tue esigenze lavorative	
Amazon Lightsail	Applicazione Linux e Windows preconfigurata per l'esecuzione di piccoli carichi di lavoro	Applicazioni Web semplici, sito Web personalizzato

- Valuta i costi (come la tariffa oraria o il trasferimento dei dati) e il sovraccarico di gestione (come l'applicazione di patch e il dimensionamento) associati a ciascuna opzione di elaborazione.
- Esegui esperimenti e benchmarking in un ambiente non di produzione per identificare quale opzione di elaborazione può soddisfare al meglio i requisiti del tuo carico di lavoro.
- Dopo aver sperimentato e identificato la tua nuova soluzione di calcolo, pianifica la migrazione e convalida i parametri prestazionali.
- Utilizza strumenti di AWS monitoraggio come Amazon CloudWatch e servizi di ottimizzazione AWS Compute Optimizerper ottimizzare continuamente le risorse di elaborazione in base a modelli di utilizzo reali.

Risorse

Documenti correlati:

- Elaborazione in cloud con AWS
- Tipi di EC2 istanze Amazon
- Amazon EKS Containers: Amazon EKS Worker Nodes
- Amazon ECS Containers: istanze di Amazon ECS Container
- Funzioni: configurazione della funzione Lambda
- Prescriptive Guidance for Containers
- Prescriptive Guidance for Serverless

Video correlati:

- AWS re:Invent 2023 - AWS Graviton: il miglior rapporto prezzo/prestazioni per i tuoi carichi di lavoro AWS
- AWS re:Invent 2023 - Nuove funzionalità di intelligenza artificiale generativa di Amazon Elastic Compute Cloud in AMS
- AWS re:Invent 2023 - What's new with Amazon Elastic Compute Cloud
- AWS re:Invent 2023 - Smart savings: Amazon Elastic Compute Cloud cost-optimization strategies
- AWS re:Invent 2021 - Powering next-gen Amazon Elastic Compute Cloud: Deep dive on the Nitro System
- AWS re:Invent 2019 - Ottimizza prestazioni e costi per le tue risorse di calcolo AWS
- AWS re:Invent 2019 - Amazon Elastic Compute Cloud foundations
- AWS re:Invent 2022 - Implementa modelli ML per l'inferenza ad alte prestazioni e a basso costo
- AWS re:Invent 2019 - Ottimizza le prestazioni e i costi per il tuo calcolo AWS
- EC2Fondazioni Amazon
- Deploy ML models for inference at high performance and low cost

Esempi correlati:

- Migrating the Web application to containers
- Esecuzione di un "Hello, World!" serverless
- EKSWorkshop Amazon
- EC2Workshop Amazon
- Efficient and Resilient Workloads with Amazon Elastic Compute Cloud Auto Scaling
- Migrazione a AWS Graviton con Container Services

PERF02-BP02 Comprendi la configurazione e le funzionalità di elaborazione disponibili

Comprendi le opzioni e le funzionalità di configurazione disponibili per il tuo servizio di calcolo in modo da fornire la giusta quantità di risorse e migliorare l'efficienza delle prestazioni.

Anti-pattern comuni:

- Non valuti le opzioni di calcolo o le famiglie di istanze disponibili rispetto alle caratteristiche del carico di lavoro.

- Esegui il provisioning eccessivo delle risorse di calcolo per soddisfare i requisiti di picco della domanda.

Vantaggi derivanti dall'adozione di questa best practice: acquisisci familiarità con le funzionalità e le configurazioni di AWS elaborazione in modo da poter utilizzare una soluzione di elaborazione ottimizzata per soddisfare le caratteristiche e le esigenze del carico di lavoro.

Livello di rischio associato se questa best practice non fosse adottata: medio

Guida all'implementazione

Ogni soluzione di calcolo ha disponibili configurazioni e funzionalità specifiche per supportare caratteristiche e requisiti diversi del carico di lavoro. Scopri in che modo puoi completare al meglio il tuo carico di lavoro e quali opzioni di configurazione sono le migliori per la tua applicazione. Esempi di queste opzioni includono famiglia di istanze, dimensioni, caratteristiche (, I/O)GPU, bursting, timeout, dimensioni delle funzioni, istanze di container e concorrenza. Se il tuo carico di lavoro utilizza la stessa opzione di elaborazione da più di quattro settimane e prevedi che le caratteristiche rimarranno le stesse in futuro, puoi usarla AWS Compute Optimizerper scoprire se l'opzione di elaborazione attuale è adatta ai carichi di lavoro e dal punto di vista della memoria. CPU

Passaggi dell'implementazione

- Comprendi i requisiti del carico di lavoro (come CPU necessità, memoria e latenza).
- AWS Consulta la documentazione e le best practice per conoscere le opzioni di configurazione consigliate che possono contribuire a migliorare le prestazioni di elaborazione. Ecco alcune opzioni di configurazione chiave da considerare:

Opzione di configurazione	Esempi
Tipo di istanza	• Le istanze ottimizzate per il calcolo sono ideali per i carichi di lavoro che richiedon o un rapporto v/memoria elevato e più elevato. CPU • Le istanze ottimizzate per la memoria offrono grandi quantità di memoria per carichi di lavoro intensivi in questo senso.

Opzione di configurazione	Esempi
	• Le istanze ottimizzate per lo storage sono progettate per carichi di lavoro che richiedon o un accesso sequenziale elevato in lettura e scrittura () allo storage locale. IOPS
Modello tariffario	• Le istanza on demand ti consentono di utilizzare la capacità di calcolo su base oraria o al secondo, senza impegni a lungo termine e sono ideali per il bursting oltre le esigenze di base per le prestazioni. • Savings Plans offrono risparmi significativi rispetto alle istanze on demand in cambio dell'impegno a utilizzare una quantità specifica di potenza di elaborazione per un periodo di uno o tre anni. • Le istanze spot ti consentono di sfruttare la capacità inutilizzata delle istanze con uno sconto per i carichi di lavoro stateless e tolleranti ai guasti.
Auto Scaling	Usa la configurazione Auto Scaling per abbinare le risorse di calcolo ai modelli di traffico.
Dimensionamento	• Usa Compute Optimizer per ricevere un efficace suggerimento di machine learning riguardo alla configurazione più adatta alle tue caratteristiche di elaborazione. • Usa AWS Lambda Power Tuning per selezionare la configurazione migliore per la tua funzione Lambda.

Opzione di configurazione	Esempi
Acceleratori di calcolo basati su hardware	• Le istanze di elaborazione accelerata eseguono funzioni come l'elaborazione grafica o la corrispondenza dei modelli di dati in modo più efficiente rispetto alle alternative basate su base. CPU • Per i carichi di lavoro di machine learning, sfrutta l'hardware appositamente progettato e specifico per il tuo carico di lavoro, come AWS Trainium, Inferentia e Amazon AWS EC2 DL1

Risorse

Documenti correlati:

- Elaborazione in cloud con AWS
- Tipi di EC2 istanze Amazon
- Controllo dello stato del processore per la tua EC2 istanza Amazon
- Amazon EKS Containers: Amazon EKS Worker Nodes
- Amazon ECS Containers: istanze di Amazon ECS Container
- Funzioni: configurazione della funzione Lambda

Video correlati:

- AWS re:Invent 2023 — AWS Graviton: il miglior rapporto prezzo/prestazioni per i tuoi carichi di lavoro AWS
- AWS re:Invent 2023 — Nuove funzionalità di intelligenza artificiale EC2 generativa di Amazon in AWS Management Console
- AWS re:Invent 2023 — Cosa c'è di nuovo con Amazon EC2
- AWS re:Invent 2023 — Risparmio intelligente: strategie di ottimizzazione dei costi di Amazon EC2
- AWS re:Invent 2021 — Potenziamento della nuova generazione di EC2 Amazon: Deep dive on the Nitro System

- AWS re:Invent 2019 — Fondamenti Amazon EC2
- AWS re:Invent 2022 — Ottimizzazione di Amazon EKS per prestazioni e costi AWS

Esempi correlati:

- Codice dimostrativo di Compute Optimizer
- Workshop sulle istanze EC2 spot di Amazon
- Carichi di lavoro efficienti e resilienti con Amazon EC2 AWS Auto Scaling
- Workshop per sviluppatori Graviton
- AWS per la giornata di immersione dei carichi di lavoro Microsoft
- AWS per una giornata di immersione nei carichi di lavoro Linux
- AWS Compute Optimizer Codice dimostrativo
- EKSWorkshop Amazon

PERF02-BP03 Raccogli metriche relative al calcolo

Registra e monitora i parametri relativi all'elaborazione per comprendere meglio le prestazioni delle tue risorse di elaborazione e migliorarne le prestazioni e l'utilizzo.

Anti-pattern comuni:

- Utilizzi solo i file di log manuali per la ricerca dei parametri.
- Utilizzi solo i parametri predefiniti registrati dal software di monitoraggio.
- Revisione dei parametri solo quando c'è un problema.

Vantaggi dell'adozione di questa best practice: la raccolta dei parametri relativi alle prestazioni ti aiuta ad allineare le prestazioni delle applicazioni ai requisiti aziendali per garantire il rispetto delle esigenze dei carichi di lavoro. Può anche aiutarti a migliorare costantemente le prestazioni e l'utilizzo delle risorse del tuo carico di lavoro.

Livello di rischio associato se questa best practice non fosse adottata: elevato

Guida all'implementazione

I carichi di lavoro del cloud possono generare grandi volumi di dati quali parametri, log ed eventi. Nel Cloud AWS, la raccolta delle metriche è un passaggio fondamentale per migliorare la sicurezza,

l'efficienza dei costi, le prestazioni e la sostenibilità. AWS fornisce un'ampia gamma di metriche relative alle prestazioni utilizzando servizi di monitoraggio come Amazon CloudWatch per fornirti informazioni preziose. Metriche come CPU l'utilizzo, l'utilizzo della memoria, l'I/O del disco e la rete in entrata e in uscita possono fornire informazioni sui livelli di utilizzo o sui colli di bottiglia delle prestazioni. Utilizza tali parametri come parte di un approccio basato sui dati per ottimizzare e ottimizzare le risorse del tuo carico di lavoro. L'ideale sarebbe raccogliere tutti i parametri relativi alle tue risorse di elaborazione in un'unica piattaforma con policy di conservazione implementate per supportare costi e obiettivi operativi.

Passaggi dell'implementazione

- Identifica quali parametri relativi alle prestazioni sono rilevanti per il tuo carico di lavoro. Raccogli i parametri sull'utilizzo delle risorse e sul modo in cui opera il tuo carico di lavoro nel cloud (come il tempo di risposta e il throughput).
  - Metriche EC2 predefinite di Amazon
  - Metriche ECS predefinite di Amazon
  - Metriche EKS predefinite di Amazon
  - Parametri predefiniti di Lambda
  - Parametri EC2 della memoria e del disco di Amazon
- Scegli e configura la soluzione di registrazione e monitoraggio giusta per il tuo carico di lavoro.
  - AWS native Observability
  - AWS Distro per OpenTelemetry
  - Amazon Managed Service per Prometheus
- Definisci il filtro e l'aggregazione richiesti per i parametri in base ai requisiti del tuo carico di lavoro.
  - Quantifica i parametri delle applicazioni personalizzate con Amazon CloudWatch Logs e filtri metrici
  - Raccogli metriche personalizzate con il tagging CloudWatch strategico di Amazon
- Configura le policy di conservazione dei dati per i parametri in modo che corrispondano ai tuoi obiettivi operativi e di sicurezza.
  - Conservazione dei dati predefinita per le metriche CloudWatch
  - Conservazione dei dati predefinita per i registri CloudWatch
- Se necessario, crea allarmi e notifiche per i parametri in modo da rispondere in modo proattivo ai problemi relativi alle prestazioni.

  - Crea allarmi per metriche personalizzate utilizzando il rilevamento delle anomalie di Amazon CloudWatch
  - Crea metriche e allarmi per pagine Web specifiche con Amazon CloudWatch RUM
- Usa l'automazione per implementare gli agenti di aggregazione di parametri e log.
  - AWS Systems Manager automazione
  - OpenTelemetryCollezionista

Risorse

Documenti correlati:

- Monitoraggio e osservabilità
- Migliori pratiche: implementazione dell'osservabilità con AWS
- CloudWatch Documentazione Amazon
- Raccogli metriche e log EC2 dalle istanze Amazon e dai server locali con l'agente CloudWatch
- Accesso ad Amazon CloudWatch Logs per AWS Lambda
- Utilizzo dei CloudWatch log con istanze di container
- Publish custom metrics
- AWS Answers: Centralized Logging
- AWS Servizi che pubblicano metriche CloudWatch
- Monitoraggio di Amazon EKS su AWS Fargate

Video correlati:

- AWS re:Invent 2023 — [LAUNCH] Monitoraggio delle applicazioni per carichi di lavoro moderni
- AWS re:Invent 2023 — Implementazione dell'osservabilità delle applicazioni
- AWS re:Invent 2023 — Creazione di una strategia di osservabilità efficace
- AWS re:Invent 2023 — Osservabilità senza interruzioni con Distro per AWS OpenTelemetry
- Gestione delle prestazioni delle applicazioni su AWS

Esempi correlati:

- AWS per Linux Workload Immersion Day- Amazon CloudWatch

- Monitoraggio di ECS cluster e container Amazon
- Monitoraggio con CloudWatch dashboard Amazon
- EKSWorkshop Amazon

PERF02-BP04 Configurazione e dimensionamento corretto delle risorse di elaborazione

Configura e dimensiona correttamente le risorse di elaborazione per soddisfare i requisiti di prestazioni del carico di lavoro ed evitare un utilizzo insufficiente o eccessivo delle risorse.

Anti-pattern comuni:

- Ignori i requisiti di prestazioni del carico di lavoro, con il risultato del provisioning eccessivo o insufficiente delle risorse di elaborazione.
- Scegli semplicemente l'istanza più grande o più piccola disponibile per tutti i carichi di lavoro.
- Usi una sola famiglia di istanze per semplificare la gestione.
- Ignori i consigli di Compute AWS Cost Explorer Optimizer o di Compute Optimizer per il corretto dimensionamento.
- Non rivaluti il carico di lavoro in base all'idoneità dei nuovi tipi di istanza.
- Certifichi solo un numero limitato di configurazioni di istanza per l'organizzazione.

Vantaggi dell'adozione di questa best practice il corretto dimensionamento delle risorse di elaborazione garantisce un funzionamento ottimale nel cloud evitando il provisioning eccessivo o insufficiente delle risorse. Il corretto dimensionamento delle risorse di elaborazione comporta in genere prestazioni ottimali e una migliore esperienza cliente, riducendo al contempo i costi.

Livello di rischio associato se questa best practice non fosse adottata: medio

Guida all'implementazione

Il dimensionamento corretto consente alle organizzazioni di gestire la propria infrastruttura cloud in modo efficiente ed economico, rispettando al contempo le esigenze aziendali. L'eccessivo provisioning delle risorse cloud può comportare costi aggiuntivi, mentre un approvvigionamento insufficiente può comportare prestazioni scadenti e un'esperienza negativa per il cliente. AWS fornisce strumenti come AWS Compute Optimizere AWS Trusted Advisorche utilizzano dati storici per fornire consigli sul corretto dimensionamento delle risorse di elaborazione.

Passaggi dell'implementazione

- Scegli il tipo di istanza più adatto alle tue esigenze:
  - Come faccio a scegliere il tipo di EC2 istanza Amazon appropriato per il mio carico di lavoro?
  - Selezione del tipo di istanza basata sugli attributi per Amazon Fleet EC2
  - Create an Auto Scaling group using attribute-based instance type selection
  - Optimizing your Kubernetes compute costs with Karpenter consolidation
- Analizza le varie caratteristiche prestazionali del tuo carico di lavoro e come queste caratteristiche si relazionano alla memoria, alla rete e all'utilizzo. CPU Utilizza questi dati per scegliere le risorse che meglio corrispondono al profilo del tuo carico di lavoro e agli obiettivi di prestazioni.
- Monitora l'utilizzo delle risorse utilizzando strumenti di AWS monitoraggio come Amazon CloudWatch.
- Seleziona la configurazione corretta per la risorsa di elaborazione.
  - Per carichi di lavoro temporanei, valuta i CloudWatch parametri di Amazon dell'istanza, ad esempio `CPUUtilization` per identificare se l'istanza è sottoutilizzata o sovrautilizzata.
  - Per carichi di lavoro stabili, controlla gli strumenti di AWS corretto dimensionamento, ad esempio e a intervalli regolari, per identificare le opportunità di ottimizzazione e dimensionamento corretto della risorsa di elaborazione AWS Compute Optimizer. AWS Trusted Advisor
- Esegui il test delle modifiche apportate alla configurazione in un ambiente non di produzione prima di implementarle in un ambiente live.
- Rivaluta costantemente nuove offerte di elaborazione e confrontale con le esigenze del carico di lavoro.

Risorse

Documenti correlati:

- Cloud Compute con AWS
- Tipi di EC2 istanze Amazon
- Amazon ECS Containers: istanze di Amazon ECS Container
- Amazon EKS Containers: Amazon EKS Worker Nodes
- Funzioni: configurazione della funzione Lambda
- Controllo dello stato del processore per la tua EC2 istanza Amazon

Video correlati:

- EC2Fondazioni Amazon
- AWS re:Invent 2023 — AWS Graviton: il miglior rapporto prezzo/prestazioni per i tuoi carichi di lavoro AWS
- AWS re:Invent 2023 — Nuove funzionalità di intelligenza artificiale EC2 generativa di Amazon in AWS Management Console
- AWS re:Invent 2023 — Cosa c'è di nuovo con Amazon EC2
- AWS re:Invent 2023 — Risparmio intelligente: strategie di ottimizzazione dei costi di Amazon EC2
- AWS re:Invent 2021 — Potenziamento della nuova generazione di EC2 Amazon: Deep dive on the Nitro System
- AWS re:Invent 2019 — Fondamenti Amazon EC2

Esempi correlati:

- AWS Compute Optimizer Codice demo
- EKSOfficina Amazon
- Right-sizing recommendations

PERF02-BP05 Scala le tue risorse di elaborazione in modo dinamico

Sfrutta l'elasticità del cloud per scalare dinamicamente le risorse di elaborazione per soddisfare le tue esigenze ed evitare un provisioning eccessivo o insufficiente per il tuo carico di lavoro.

Anti-pattern comuni:

- Risposta agli allarmi aumentando manualmente la capacità.
- Utilizzi le stesse linee guida per il dimensionamento (generalmente infrastruttura statica) di quelle on-premises.
- Dopo un evento di dimensionamento, lasci una capacità aumentata anziché ridurre il dimensionamento.

Vantaggi dell'adozione di questa best practice: la configurazione e il test dell'elasticità delle risorse di elaborazione possono aiutarti a risparmiare denaro, mantenere i benchmark delle prestazioni e migliorare l'affidabilità al variare del traffico.

Livello di rischio associato se questa best practice non fosse adottata: elevato

Guida all'implementazione

AWS offre la flessibilità necessaria per aumentare o ridurre le risorse in modo dinamico attraverso una varietà di meccanismi di scalabilità al fine di soddisfare le variazioni della domanda. In combinazione con i parametri relativi all'elaborazione, il dimensionamento dinamico consente ai carichi di lavoro di rispondere automaticamente alle modifiche e utilizzare il set ottimale di risorse di elaborazione per raggiungere l'obiettivo.

Puoi adottare varie strategie di approccio per associare l'offerta di risorse alla domanda.

- Approccio al tracciamento degli obiettivi: monitora il parametro di dimensionamento e aumenta o diminuisci automaticamente la capacità in base alle esigenze.
- Dimensionamento predittivo: procedi a ridurre orizzontalmente in previsione delle tendenze giornaliere e settimanali.
- Approccio basato sulla pianificazione: imposta il tuo programma di dimensionamento in base alle variazioni di carico prevedibili.
- Scalabilità del servizio: scegli i servizi (come quelli serverless) che si dimensionano automaticamente per progettazione.

Assicurati che le implementazioni dei carichi di lavoro siano in grado di gestire eventi che prevedono l'aumentare verticalmente e il ridurre verticalmente.

Passaggi dell'implementazione

- Istanze di elaborazione, container e funzioni forniscono tutti meccanismi di elasticità, in combinazione con il dimensionamento automatico o sotto forma di funzionalità del servizio. Ecco alcuni esempi di meccanismi di dimensionamento automatico:

Meccanismo di scalabilità automatica	Dove usarlo
Amazon EC2 Auto Scaling	Per assicurarti di disporre del numero corretto di EC2 istanze Amazon disponibili per gestire il carico di utenti per la tua applicazione.
Application Auto Scaling	Per scalare automaticamente le risorse per singoli AWS servizi oltre Amazon, EC2 come

Meccanismo di scalabilità automatica	Dove usarlo
	AWS Lambdafunzioni o servizi Amazon Elastic Container Service (AmazonECS).
Kubernetes Cluster Autoscaler/Karpenter	Dimensiona automaticamente i cluster Kubernetes.

- La scalabilità viene spesso discussa in relazione a servizi di elaborazione come Amazon EC2 Instances o funzioni. AWS Lambda Assicurati di considerare anche la configurazione di servizi non di calcolo come AWS Glue per soddisfare la domanda.
- Verifica che i parametri per il dimensionamento corrispondano alle caratteristiche del carico di lavoro da implementare. Se stai implementando un'applicazione di transcodifica video, è previsto un CPU utilizzo del 100% e non dovrebbe essere il parametro principale. Utilizza la profondità della coda dei processi di transcodifica. Se necessario, puoi utilizzare una metrica personalizzata per la tua policy di dimensionamento. Per scegliere le metriche giuste, consulta le seguenti linee guida per AmazonEC2:
  - La metrica deve essere una metrica di utilizzo valida e descrivere il livello di impiego di un'istanza.
  - Il valore del parametro deve aumentare e diminuire in proporzione al numero di istanze nel gruppo con scalabilità automatica.
- Assicurati di utilizzare il dimensionamento dinamico anziché il dimensionamento manuale per il tuo gruppo Auto Scaling. È consigliabile utilizzare le policy di dimensionamento del monitoraggio degli obiettivi nel dimensionamento dinamico
- Verifica che le implementazioni dei carichi di lavoro siano in grado di gestire entrambi gli eventi di dimensionamento (aumento e riduzione). Ad esempio, puoi usare la cronologia delle attività per verificare le attività di ridimensionamento per un gruppo Auto Scaling.
- Analizza il tuo carico di lavoro per individuare modelli prevedibili e dimensionare le tue risorse in modo proattivo, anticipando variazioni nella domanda previste e pianificate. Con il dimensionamento predittivo puoi eliminare la necessità di offrire capacità in eccedenza. Per maggiori dettagli, consulta Predictive Scaling with Amazon Auto EC2 Scaling.

Risorse

Documenti correlati:

- Cloud Compute con AWS

- Tipi di EC2 istanze Amazon
- Amazon ECS Containers: istanze di Amazon ECS Container
- Amazon EKS Containers: Amazon EKS Worker Nodes
- Funzioni: configurazione della funzione Lambda
- Controllo dello stato del processore per la tua EC2 istanza Amazon
- Approfondimento su Amazon ECS Cluster Auto Scaling
- Introducing Karpenter – An Open-Source High-Performance Kubernetes Cluster Autoscaler

Video correlati:

- AWS re:Invent 2023 — AWS Graviton: il miglior rapporto prezzo/prestazioni per i tuoi carichi di lavoro AWS
- AWS re:Invent 2023 — Nuove funzionalità di intelligenza artificiale EC2 generativa di Amazon nella console di gestione AWS
- AWS re:Invent 2023 — Cosa c'è di nuovo con Amazon EC2
- AWS re:Invent 2023 — Risparmio intelligente: strategie di ottimizzazione dei costi di Amazon EC2
- AWS re:Invent 2021 — Potenziamento della nuova generazione di EC2 Amazon: Deep dive on the Nitro System
- AWS re:Invent 2019 — Fondamenti Amazon EC2

Esempi correlati:

- Esempi di EC2 gruppi Amazon Auto Scaling
- EKSWorkshop Amazon
- Scala i tuoi EKS carichi di lavoro Amazon eseguendoli su IPv6

PERF02-BP06 Usa acceleratori di calcolo ottimizzati basati su hardware

Utilizza gli acceleratori hardware per eseguire determinate funzioni in modo più efficiente rispetto alle alternative CPU basate.

Anti-pattern comuni:

- Nel carico di lavoro non hai confrontato un'istanza per uso generico con un'istanza dedicata in grado di offrire prestazioni più elevate e costi inferiori.

- Stai utilizzando acceleratori di calcolo basati su hardware per attività che possono essere più efficienti utilizzando alternative basate su hardware. CPU
- Non stai monitorando l'utilizzo. GPU

Vantaggi derivanti dall'adozione di questa best practice: utilizzando acceleratori basati su hardware, come le unità di elaborazione grafica (GPUs) e gli array di porte programmabili sul campo (FPGAs), è possibile eseguire determinate funzioni di elaborazione in modo più efficiente.

Livello di rischio associato se questa best practice non fosse adottata: medio

Guida all'implementazione

Le istanze di elaborazione accelerata forniscono l'accesso ad acceleratori di calcolo basati su hardware come and. GPUs FPGAs Questi acceleratori hardware eseguono determinate funzioni, come l'elaborazione grafica o la corrispondenza dei modelli di dati, in modo più efficiente rispetto alle alternative basate. CPU Molti carichi di lavoro accelerati, come il rendering grafico, la transcodifica e il machine learning, sono altamente variabili in termini di utilizzo di risorse. Esegui questo hardware solo per il tempo necessario e disattivalo con l'automazione quando non serve per migliorare l'efficienza complessiva delle prestazioni.

Passaggi dell'implementazione

- Identifica le istanza a calcolo accelerato in grado di soddisfare i tuoi requisiti.
- Per i carichi di lavoro di machine learning, sfrutta l'hardware appositamente progettato e specifico per il tuo carico di lavoro, come AWS Trainium, Inferentia e Amazon.AWS EC2 DL1 AWS Le istanze Inferentia come le istanze Inf2 offrono prestazioni/watt migliori fino al 50% rispetto alle istanze Amazon comparabili. EC2
- Raccogli i parametri di utilizzo delle istanze a calcolo accelerato. Ad esempio, puoi utilizzare l' CloudWatch agente per raccogliere metriche come `utilization_gpu` e `utilization_memory` per le tue, GPUs come mostrato in Collect NVIDIA GPU metrics with Amazon. CloudWatch
- Ottimizza il codice, il funzionamento della rete e le impostazioni degli acceleratori hardware per garantire il pieno utilizzo dell'hardware sottostante.
  - Ottimizza le impostazioni GPU
  - GPUMonitoraggio e ottimizzazione nel deep learning AMI
  - Ottimizzazione dell'I/O per l'ottimizzazione GPU delle prestazioni della formazione di deep learning in Amazon SageMaker

- Utilizza le librerie e i driver più recenti ad alte prestazioni. GPU
- Usa l'automazione per rilasciare GPU istanze quando non sono in uso.

Risorse

Documenti correlati:

- Utilizzo di GPUs Amazon Elastic Container Service
- GPUistanze
- Istanze con Trainium AWS
- Istanze con Inferentia AWS
- Let's Architect! Architecting with custom chips and accelerators

- Calcolo accelerato
- EC2VT1Istanze Amazon
- Come faccio a scegliere il tipo di EC2 istanza Amazon appropriato per il mio carico di lavoro?
- Scegli l'acceleratore di intelligenza artificiale e la compilazione di modelli migliori per l'inferenza della visione artificiale con Amazon SageMaker

Video correlati:

- AWS re:Invent 2021 - Come selezionare le istanze Amazon Elastic Compute Cloud GPU per il deep learning
- AWS re:Invent 2022 - [!] NEW LAUNCH Presentazione delle AWS istanze Amazon Inf2 basate su Inferentia2 EC2
- AWS re:Invent 2022 - Accelerate deep learning and innovate faster with AWS Trainium
- AWS re:Invent 2022 - Apprendimento approfondito con: dalla formazione alla distribuzione AWS NVIDIA

Esempi correlati:

- Amazon SageMaker e il NVIDIA GPU cloud (NGC)
- Utilizzalo SageMaker con Trainium e Inferentia per carichi di lavoro ottimizzati di deep learning, training e inferenza

- Ottimizzazione dei NLP modelli con istanze Amazon Elastic Compute Cloud Inf1 in Amazon SageMaker

Gestione dei dati

Questions

- PERF3. In che modo archivi, gestisci e accedi ai dati nel tuo carico di lavoro?

PERF3. In che modo archivi, gestisci e accedi ai dati nel tuo carico di lavoro?

La soluzione di gestione dei dati ottimale per un particolare sistema varia in base al tipo di dati (blocco, file o oggetto), ai modelli di accesso (casuale o sequenziale), alla velocità effettiva richiesta, alla frequenza di accesso (online, offline, di archiviazione), alla frequenza di aggiornamento (WORMdinamica) e ai vincoli di disponibilità e durabilità. I carichi di lavoro Well-Architected utilizzano archivi dati appositamente progettati che impiegano diverse funzionalità per migliorare le prestazioni.

Best practice

- PERF03-BP01 Utilizza un archivio dati appositamente progettato che supporti al meglio i requisiti di accesso e archiviazione dei dati
- PERF03-BP02 Valuta le opzioni di configurazione disponibili per l'archivio dati
- PERF03-BP03 Raccolta e registrazione delle metriche delle prestazioni degli archivi dati
- PERF03-BP04 Implementazione di strategie per migliorare le prestazioni delle query nell'archivio dati
- PERF03-BP05 Implementare modelli di accesso ai dati che utilizzano la memorizzazione nella cache

PERF03-BP01 Utilizza un archivio dati appositamente progettato che supporti al meglio i requisiti di accesso e archiviazione dei dati

Comprendi le caratteristiche dei dati (come la condivisione, le dimensioni, la dimensione della cache, gli schemi di accesso, la latenza, il throughput e la persistenza dei dati) per selezionare i data store (archiviazione o database) dedicati per il tuo carico di lavoro.

Anti-pattern comuni:

- Continui a utilizzare un datastore per via dell'esperienza e delle competenze interne relative a quel particolare tipo di soluzione di database.
- Ritieni che tutti i carichi di lavoro abbiano requisiti di accesso e archiviazione di dati simili.
- Non hai implementato un catalogo di dati per eseguire l'inventario dei tuoi asset.

Vantaggi dell'adozione di questa best practice: la comprensione delle caratteristiche e dei requisiti dei dati ti consente di determinare la tecnologia di archiviazione più efficiente e performante appropriata per le tue esigenze del carico di lavoro.

Livello di rischio associato se questa best practice non fosse adottata: elevato

Guida all'implementazione

Quando selezionate e implementate l'archiviazione dei dati, assicuratevi che le caratteristiche di interrogazione, scalabilità e archiviazione supportino i requisiti relativi ai dati del carico di lavoro. AWS offre numerose tecnologie di archiviazione dei dati e di database, tra cui storage a blocchi, storage di oggetti, storage in streaming, file system, database relazionali, chiave-valore, documentali, in memoria, grafici, di serie temporali e di registro. Ogni soluzione di gestione dei dati offre soluzioni e configurazioni adatte a gestire i tuoi casi d'uso e modelli di dati. Comprendendo le caratteristiche e i requisiti dei dati, è possibile abbandonare la tecnologia di storage monolitica e adottare approcci restrittivi per concentrarsi sulla gestione appropriata dei dati. one-size-fits-all

Passaggi dell'implementazione

- Esegui un inventario dei vari tipi di dati esistenti nel tuo carico di lavoro.
- Comprendi e documenta le caratteristiche e i requisiti dei dati, tra cui:
  - Tipo di dati (non strutturati, semi-strutturati, relazionali)
  - Volume e crescita dei dati
  - Durabilità dei dati: persistenti, effimeri, transitori
  - ACIDrequisiti (atomicità, coerenza, isolamento, durabilità)
  - Schemi di accesso ai dati (con uso intensivo di lettura o scrittura)
  - Latenza
  - Prestazioni
  - IOPS(operazioni di ingresso/uscita al secondo)
  - Periodo di conservazione dei dati

- Scopri i diversi archivi di dati (servizi di archiviazione e database) disponibili per il tuo carico di lavoro e AWS che possono soddisfare le caratteristiche dei tuoi dati, come descritto in. PERF01-BP01 Scopri e comprendi i servizi e le funzionalità cloud disponibili Alcuni esempi di tecnologie di archiviazione AWS e delle loro caratteristiche chiave sono:

Tipo	AWS Servizi	Caratteristiche chiave
Archiviazione di oggetti	Amazon S3	Scalabilità illimitata, alta disponibilità e molteplic i opzioni di accessibilità. L'accesso a oggetti e il relativo trasferimento da e verso Amazon S3 può utilizzare un servizio, come Transfer Accelerat ion o Punti di accesso, per supportare la posizione, le esigenze di sicurezza e i modelli di accesso.
Archiviazione	Amazon S3 Glacier	Progettato per l'archiviazione dei dati.
Archiviazione in streaming	Amazon Kinesis Streaming gestito da Amazon per Apache Kafka (Amazon) MSK	Acquisizione e archiviazione efficienti dei dati in streaming.
File system condiviso	Amazon Elastic File System (AmazonEFS)	File system montabile a cui è possibile accedere da più tipi di soluzioni di calcolo.

Tipo	AWS Servizi	Caratteristiche chiave
File system condiviso	Amazon FSx	Basato sulle più recenti soluzioni di AWS elaborazione per supportare quattro file system di uso comune: Open NetApp ONTAPZFS, Windows File Server e Lustre. FSxLa latenza, la velocità effettiva e la velocità effettiva di Amazon IOPS variano in base al file system e devono essere prese in considerazione quando si seleziona il file system giusto per le esigenze di carico di lavoro.
Storage a blocchi	Amazon Elastic Block Store (AmazonEBS)	Servizio di storage a blocchi scalabile e ad alte prestazioni progettato per Amazon Elastic Compute Cloud (Amazon). EC2 Amazon EBS include storage SSD supportato per carichi di lavoro transazionali e intensivi e HDD storage supportato per carichi di lavoro con throughput IOPS intensivo.

Tipo	AWS Servizi	Caratteristiche chiave
Database relazionale	Amazon Aurora, AmazonRDS, Amazon Redshift.	Progettato per supportare transazioni ACID (atomicità, coerenza, isolamento, durabilità) e mantenere l'integrità referenziale e una forte coerenza dei dati. Molte applicazioni tradizionali, la pianificazione delle risorse aziendali (ERP), la gestione delle relazioni con i clienti (CRM) e l'e-commerce utilizzano database relazionali per archiviare i propri dati.
Database chiave-valore	Amazon DynamoDB	Ottimizzato per schemi di accesso di uso comune, in genere per archiviare e recuperare grandi volumi di dati. Le app Web dal traffico elevato, i sistemi di e-commerce e le applicazioni di videogiochi sono casi d'uso tipici dei database chiave-valore.
Database di documenti	Amazon DocumentDB	Progettato per archiviare dati semistrutturati come documenti similiJSON. Questi database aiutano gli sviluppatori a creare e aggiornare rapidamente applicazioni quali gestione di contenuti, cataloghi e profili utente.

Tipo	AWS Servizi	Caratteristiche chiave
Database in memoria	Amazon ElastiCache, Amazon MemoryDB per Redis	Vengono utilizzati per applicazioni che richiedon o accesso in tempo reale ai dati, bassissima latenza ed elevatissimo throughput. È possibile utilizzare database in memoria per la memorizza zione nella cache delle applicazioni, la gestione delle sessioni, la classifica dei giochi, l'archivio delle caratteri stiche ML a bassa latenza, il sistema di messaggistica dei microservizi e un meccanism o di streaming a elevato throughput.
Database a grafo	Amazon Neptune	Utilizzato con le applicazi oni che devono navigare ed eseguire query su milioni di relazioni tra set di dati a grafo altamente connessi, con una latenza misurata in milliseco ndi su larga scala. Molte aziende utilizzano database a grafo per il rilevamen to di attività fraudolente, i social network e i motori di raccomandazione.

Tipo	AWS Servizi	Caratteristiche chiave
Database di serie temporali	Amazon Timestream	Utilizzato per raccogliere, sintetizzare e derivare in modo efficiente approfond imenti dai dati che cambiano nel tempo. Le applicazioni IoT e DevOps la telemetria industriale possono utilizzare database di serie temporali.
Colonna ampia	Amazon Keyspaces (per Apache Cassandra)	Utilizza tabelle, righe e colonne, ma a differenza di un database relaziona le, i nomi e il formato delle colonne possono variare da riga a riga all'interno della stessa tabella. In genere, gli store colonnari sono utilizzati nelle applicazioni industriali su larga scala per la manutenzione delle apparecchiature, la gestione delle flotte e l'ottimizzazione dei percorsi.
Di libri mastri	Database Amazon Quantum Ledger (Amazon) QLDB	Fornisce un'autorità centralizzata e affidabile per mantenere un registro delle transazioni scalabile , immutabile e verificabile tramite crittografia per ogni applicazione. I database di libri mastri vengono utilizzati per sistemi di record, catena di fornitura, registrazioni e persino transazioni bancarie.

- Se stai creando una piattaforma dati, sfrutta un'architettura di dati moderna AWS per integrare il tuo data lake, il data warehouse e gli archivi dati creati appositamente.
- Le domande chiave da porsi quando si sceglie un data store per il carico di lavoro sono le seguenti:

Domanda	Aspetti da considerare
Come sono strutturati i dati?	• Se i dati non sono strutturati, prendi in considerazione un object store come Amazon S3 o un database No SQL come Amazon DocumentDB • Per i dati chiave-valore, prendi in considerazione DynamoDB, Amazon (ElastiCache Redis) o Amazon MemoryDB OSS
Quale livello di integrità referenziale è richiesto ?	• Per i vincoli di chiave esterna, i database relazionali come Amazon e RDS Aurora possono fornire questo livello di integrità. • In genere, all'interno di un SQL modello senza dati, i dati vengono denormalizzati in un unico documento o in una raccolta di documenti da recuperare in un'unica richiesta anziché riunirli tra più documenti o tabelle.
È richiesta la conformità ACID (atomicità, coerenza, isolamento, durabilità)?	• Se le ACID proprietà associate ai database relazionali sono obbligatorie, prendi in considerazione un database relazionale come Amazon e RDS Aurora. • Se è richiesta una forte coerenza per No SQL database, è possibile utilizzare letture fortemente coerenti con DynamoDB.
Come cambierà nel tempo l'archiviazione? In che modo questo avrà effetto sulla scalabilità?	• I database serverless come DynamoDB e Amazon Quantum Ledger Database (Amazon) verranno scalati dinamicamente. QLDB

Domanda	Aspetti da considerare
	• Per i database relazionali sono previsti limiti massimi per l'archiviazione allocata, al raggiungimento dei quali si rende spesso necessario partizionare orizzontalmente tali database tramite meccanismi quali la partizione.
Qual è la proporzione di query in lettura rispetto alle quelle in scrittura? Il caching potrebbe probabilmente migliorare le prestazioni?	• I carichi di lavoro impegnativi in lettura possono trarre vantaggio da un livello di caching, ad esempio DAXse il database è ElastiCacheDynamoDB. • Le letture possono anche essere scaricate per leggere le repliche con database relazionali come Amazon. RDS
L'archiviazione e la modifica (OLTP- Elaborazione delle transazioni online) o il recupero e il reporting (OLAP- Elaborazione analitica online) hanno una priorità più elevata?	• Per l'elaborazione transazionale read-as-is ad alto rendimento, prendi in considerazione un database No come DynamoDB. SQL • Per modelli di lettura complessi e ad alta velocità (come join) con coerenza, usa Amazon. RDS • Per le query analitiche, prendi in considerazione un database a colonne come Amazon Redshift o l'esportazione dei dati su Amazon S3 e l'esecuzione di analisi utilizzando Athena o Amazon. QuickSight

Domanda	Aspetti da considerare
Che livello di durabilità è necessario per i dati?	• Aurora replica automaticamente i dati su tre zone di disponibilità all'interno di una regione, il che significa che i dati sono altamente durevoli con minori probabilità di perdite. • DynamoDB viene automaticamente replicato in più zone di disponibilità per offrire livelli elevati di disponibilità e durabilità dei dati. • Amazon S3 offre il 99,999999999 di durabilità. Molti servizi di database, come Amazon RDS e DynamoDB, supportano l'esportazione di dati su Amazon S3 per la conservazione e l'archiviazione a lungo termine.
È presente il desiderio di abbandonare i motori di database commerciali o i costi di licenza?	• Prendi in considerazione motori open source come Postgre SQL e My on SQL Amazon o RDS Aurora. • Sfrutta AWS Database Migration Service e AWS Schema Conversion Tool per eseguire le migrazioni dai motori di database commerciali a quelli open-source
Quali sono le aspettative operative per il database? Il passaggio ai servizi gestiti è una priorità?	• Sfruttare Amazon RDS anziché Amazon EC2 e DynamoDB o Amazon DocumentD B anziché ospitare autonomamente SQL un database No può ridurre il sovraccarico operativo.

Domanda	Aspetti da considerare
Come avviene attualmente l'accesso al database? Si tratta solo di accesso alle applicazioni o ci sono utenti di business intelligence (BI) e altre applicazioni connesse? off-the-shelf	• In presenza di dipendenze da strumenti esterni, potresti dover mantenere la compatibilità con i database che supportan o. Amazon RDS è completamente compatibi le con le diverse versioni del motore che supporta, tra cui Microsoft SQL Server, OracleSQL, My e SQL Postgre.

- Esegui esperimenti e benchmarking in un ambiente non di produzione per identificare quale datastore può soddisfare al meglio i requisiti del tuo carico di lavoro.

Risorse

Documenti correlati:

- Tipi di EBS volume Amazon
- EC2Archiviazione Amazon
- AmazonEFS: EFS Prestazioni Amazon
- Prestazioni FSx di Amazon for Lustre
- Prestazioni FSx di Amazon per Windows File Server
- Amazon S3 Glacier: documentazione di S3 Glacier
- Amazon S3: considerazioni su velocità e prestazioni delle richieste
- Archiviazione su cloud con AWS
- Caratteristiche di Amazon EBS I/O
- Database su cloud con AWS
- AWS Database Caching
- DynamoDB Accelerator
- Best practice di Amazon Aurora
- Prestazioni di Amazon RedShift
- Amazon Athena top 10 performance tips
- Amazon Redshift Spectrum best practices

- Amazon DynamoDB best practices
- Scegli tra Amazon EC2 e Amazon RDS
- Best practice per l'implementazione di Amazon ElastiCache

Video correlati:

- AWS re:Invent 2023: migliora l'efficienza di Amazon Elastic Block Store e sii più efficiente in termini di costi
- AWS re:Invent 2023: ottimizzazione del prezzo e delle prestazioni dello storage con Amazon Simple Storage Service
- AWS re:Invent 2023: creazione e ottimizzazione di un data lake su Amazon Simple Storage Service
- AWS re:Invent 2022: creazione di moderne architetture di dati su AWS
- AWS re:Invent 2022: creazione di architetture di data mesh su AWS
- AWS re:Invent 2023: approfondimenti su Amazon Aurora e sulle sue innovazioni
- AWS re:Invent 2023: modellazione avanzata dei dati con Amazon DynamoDB
- AWS re:Invent 2022: modernizza le app con database creati appositamente
- Amazon DynamoDB deep dive: Advanced design patterns

Esempi correlati:

- AWS Workshop sui database creati appositamente
- Databases for Developers
- AWS Giornata di immersione nell'architettura dei dati moderna
- Crea una rete di dati su AWS
- Esempi di Amazon S3
- Optimize Data Pattern using Amazon Redshift Data Sharing
- Migrazioni dei database
- MS SQL Server - AWS Database Migration Service (AWS DMS) Demo sulla replica
- Database Modernization Hands On Workshop
- Esempi di Amazon Neptune

PERF03-BP02 Valuta le opzioni di configurazione disponibili per l'archivio dati

Comprendi e valuta le varie funzionalità e opzioni di configurazione disponibili per i tuoi datastore per ottimizzare lo spazio di archiviazione e le prestazioni per il tuo carico di lavoro.

Anti-pattern comuni:

- Utilizzi solo un tipo di storage, ad esempio AmazonEBS, per tutti i carichi di lavoro.
- Utilizzi il provisioning IOPS per tutti i carichi di lavoro senza eseguire test reali su tutti i livelli di storage.
- Non conosci le opzioni di configurazione della soluzione di gestione dei dati scelta.
- Ti basi soltanto sull'aumento delle dimensioni dell'istanza, senza tenere conto di altre opzioni di configurazione disponibili.
- Non esegui il test delle caratteristiche di dimensionamento del tuo datastore.

Vantaggi dell'adozione di questa best practice: esplorare le configurazioni del datastore e sperimentare con esse può consentire di ridurre il costo dell'infrastruttura, migliorare le prestazioni e ridurre l'impegno richiesto per mantenere i carichi di lavoro.

Livello di rischio associato se questa best practice non fosse adottata: medio

Guida all'implementazione

Un carico di lavoro può utilizzare uno o più datastore in base ai requisiti di archiviazione di dati e relativo accesso. Per ottimizzare prestazioni, efficienza e costi, è necessario valutare gli schemi di accesso ai dati per determinare le configurazioni appropriate del datastore. Nella valutazione delle opzioni di datastore, prendi in considerazione vari aspetti come le opzioni di archiviazione, la memoria, l'elaborazione, la replica di lettura, i requisiti di coerenza, il pool di connessioni e le opzioni di caching. Esegui esperimenti con queste diverse opzioni di configurazione per migliorare i parametri di efficienza delle prestazioni.

Passaggi dell'implementazione

- Esamina le configurazioni correnti (come il tipo di istanza, la dimensione di archiviazione o la versione del motore di database) del tuo datastore.
- AWS Consulta la documentazione e le best practice per conoscere le opzioni di configurazione consigliate che possono contribuire a migliorare le prestazioni del tuo data store. Le principali opzioni da considerare per il datastore sono le seguenti:

Opzione di configurazione	Esempi
Riduzione del carico delle letture (come le repliche di lettura e la memorizzazione nella cache)	• Per le tabelle DynamoDB, è possibile scaricare le letture utilizzando per la memorizzazione nella cache. DAX • Puoi creare un cluster Amazon ElastiCache (RedisOSS) e configurare l'applicazione in modo che legga prima dalla cache, per poi tornare al database se l'elemento richiesto non è presente. • I database relazionali come Amazon RDS e Aurora e i database forniti SQL No come Neptune e Amazon DocumentDB supportan o tutti l'aggiunta di repliche di lettura per scaricare le parti di lettura del carico di lavoro. • I database serverless come DynamoDB si dimensionano automaticamente. Assicurat i di disporre di unità di capacità di lettura sufficienti () per gestire il carico di lavoro. RCU

Opzione di configurazione	Esempi
Dimensionamento delle scritture (come la partizione delle chiavi di partizione o l'introduzione di una coda)	• Per i database relazionali, è possibile aumentare le dimensioni dell'istanza per far fronte a un maggiore carico di lavoro o aumentare il provisioning IOPs per consentire una maggiore velocità di trasmissione dello storage sottostante. • È anche possibile introdurre una coda davanti al database, invece di eseguire direttamente la scrittura su di esso. Questo schema consente di disaccoppiare l'acquisizione dal database e controllare il flusso, in modo che il database sia in grado di gestirlo. • Raggruppare in batch le richieste di scrittura, anziché creare molte transazioni di breve durata, può aiutare a migliorare il throughput in database relazionali con un elevato volume in scrittura. • I database serverless come DynamoDB possono scalare il throughput di scrittura automaticamente o regolando le unità WCU di capacità di scrittura assegnate () a seconda della modalità di capacità. • È tuttavia possibile che si verifichino problemi con le partizioni hot quando si raggiungono i limiti di throughput per una determinata chiave di partizione. Questo problema può essere arginato scegliendo una chiave di partizione con una distribuzione più uniforme o eseguendo lo sharding in lettura della chiave di partizione.

Opzione di configurazione	Esempi
Policy per gestire il ciclo di vita dei set di dati	• Puoi utilizzare il ciclo di vita Amazon S3 per gestire gli oggetti durante il loro ciclo di vita. In caso di schemi di accesso sconosciuti, mutevoli o imprevedibili, puoi utilizzare il Piano intelligente Amazon S3, che monitora gli schemi di accesso e sposta in automatic o gli oggetti che non hanno fatto registrar e accessi a livelli di accessi più economici . Sfrutta i parametri di Amazon S3 Storage Lens per individuare opportunità di ottimizza zione e lacune nella gestione del ciclo di vita. • Amazon EFS Lifecycle Management gestisce automaticamente lo storage dei file per i tuoi file system.
Gestione e pooling delle connessioni	• Amazon RDS Proxy può essere utilizzato con Amazon RDS e Aurora per gestire le connessioni al database. • I database serverless come DynamoDB non hanno connessioni associate, ma valuta la capacità assegnata e le policy di dimension amento automatico per affrontare i picchi nel carico.

- Esegui esperimenti e benchmarking in un ambiente non di produzione per identificare quale opzione di configurazione può soddisfare i requisiti del tuo carico di lavoro.
- Dopo gli esperimenti, pianifica la migrazione e convalida i parametri delle prestazioni.
- Utilizza strumenti di AWS monitoraggio (come Amazon CloudWatch) e ottimizzazione (come Amazon S3 Storage Lens) per ottimizzare continuamente il tuo archivio dati utilizzando modelli di utilizzo reali.

Risorse

Documenti correlati:

- Archiviazione nel cloud in AWS
- Tipi di EBS volume Amazon
- EC2Archiviazione Amazon
- AmazonEFS: EFS prestazioni Amazon
- Prestazioni Amazon FSx for Lustre
- Prestazioni FSx di Amazon per Windows File Server
- Amazon S3 Glacier: documentazione di S3 Glacier
- Amazon S3: considerazioni su velocità e prestazioni delle richieste
- Caratteristiche di Amazon EBS I/O
- Database su cloud con AWS
- AWS Database Caching
- DynamoDB Accelerator
- Best practice di Amazon Aurora
- Prestazioni di Amazon RedShift
- Amazon Athena top 10 performance tips
- Amazon Redshift Spectrum best practices
- Amazon DynamoDB best practices

Video correlati:

- AWS re:Invent 2023: Improve Amazon Elastic Block Store efficiency and be more cost-efficient
- AWS re:Invent 2023: Optimize storage price and performance with Amazon Simple Storage Service
- AWS re:Invent 2023: Building and optimizing a data lake on Amazon Simple Storage Service
- AWS re:Invent 2023: Cosa c'è di nuovo con lo storage di file AWS
- AWS re:Invent 2023: Dive deep into Amazon DynamoDB

Esempi correlati:

- AWS Workshop sui database appositamente progettati
- Databases for Developers
- AWS Giornata di immersione nell'architettura dei dati moderna

- Amazon EBS Scalabilità automatica
- Esempi di Amazon S3
- Esempi di Amazon DynamoDB
- AWS Esempi di migrazione di database
- Workshop sulla modernizzazione dei database
- Utilizzo dei parametri sul database Amazon RDS for Postgress

PERF03-BP03 Raccolta e registrazione delle metriche delle prestazioni degli archivi dati

Tieni traccia e registra i parametri delle prestazioni pertinenti per il tuo datastore per capire l'andamento delle prestazioni delle soluzioni di gestione dei dati. Questi parametri possono aiutarti a ottimizzare il tuo datastore, verificare che i requisiti del carico di lavoro siano rispettati e fornire una panoramica chiara sull'andamento delle prestazioni del carico di lavoro.

Anti-pattern comuni:

- Utilizzi solo i file di log manuali per la ricerca dei parametri.
- Pubblichi i parametri solo sugli strumenti interni utilizzati dal tuo team e non hai un quadro completo del carico di lavoro.
- Utilizzo solo dei parametri predefiniti registrati dal software di monitoraggio selezionato.
- Revisione dei parametri solo quando c'è un problema.
- Monitori solo i parametri a livello di sistema, senza acquisire i parametri di accesso ai dati o di utilizzo.

Vantaggi dell'adozione di questa best practice: la definizione di una linea di base delle prestazioni ti aiuta a comprendere il comportamento normale e i requisiti dei carichi di lavoro. Gli schemi anomali possono essere identificati ed eliminati più rapidamente, per migliorare le prestazioni e l'affidabilità del datastore.

Livello di rischio associato se questa best practice non fosse adottata: elevato

Guida all'implementazione

Per monitorare le prestazioni dei datastore, devi registrare più parametri delle prestazioni in un periodo di tempo. Ciò consente di rilevare le anomalie e di misurare le prestazioni rispetto ai parametri aziendali, per verificare che le esigenze del carico di lavoro siano rispettate.

I parametri devono includere sia il sistema sottostante che supporta il datastore sia i parametri del database. Le metriche di sistema sottostanti possono includere CPU utilizzo, memoria, spazio di archiviazione su disco disponibile, I/O su disco, rapporto di accesso alla cache e metriche di rete in entrata e in uscita, mentre le metriche del data store possono includere transazioni al secondo, query principali, tassi medi di query, tempi di risposta, utilizzo dell'indice, blocchi delle tabelle, timeout delle query e numero di connessioni aperte. Questi dati sono cruciali per capire l'andamento del carico di lavoro e come viene utilizzata la soluzione di gestione dei dati. Utilizza tali parametri come parte di un approccio basato sui dati per mettere a punto e ottimizzare le risorse del tuo carico di lavoro.

Utilizza strumenti, librerie e sistemi che registrano misure delle prestazioni relative alle prestazioni del database.

Passaggi dell'implementazione

- Determina i principali parametri delle prestazioni da monitorare per il tuo datastore.
  - Parametri e dimensioni di Amazon S3
  - Parametri di monitoraggio per un'istanza Amazon RDS
  - Monitoraggio del carico del DB con Performance Insights su Amazon RDS
  - Panoramica sul monitoraggio avanzato
  - DynamoDB Metrics and dimensions
  - Monitoraggio di DynamoDB Accelerator
  - Monitoraggio di Amazon MemoryDB con Amazon CloudWatch
  - Quali parametri è opportuno monitorare?
  - Monitoring Amazon Redshift cluster performance
  - Timestream metrics and dimensions
  - CloudWatch Parametri Amazon per Amazon Aurora
  - Creazione di log e monitoraggio in Amazon Keyspaces (per Apache Cassandra)
  - Monitoring Amazon Neptune Resources
- Utilizza una soluzione di registrazione e monitoraggio approvata per raccogliere queste metriche. Amazon CloudWatch può raccogliere metriche tra le risorse della tua architettura. Puoi anche raccogliere e pubblicare parametri personalizzati per ottenere parametri aziendali o derivati. Utilizza soluzioni CloudWatch di terze parti per impostare allarmi che indicano quando vengono superate le soglie.
- Verifica se il monitoraggio dei datastore può trarre vantaggio da una soluzione di machine learning che rileva le anomalie delle prestazioni.

  - Amazon DevOps Guru per Amazon RDS offre visibilità sui problemi di prestazioni e fornisce consigli per azioni correttive.
- Configura la conservazione dei dati nella soluzione di monitoraggio e registrazione per soddisfare i tuoi obiettivi operativi e di sicurezza.
  - Conservazione dei dati predefinita per le metriche CloudWatch
  - Conservazione dei dati predefinita per i registri CloudWatch

Risorse

Documenti correlati:

- AWS Database Caching
- Amazon Athena top 10 performance tips
- Amazon Aurora best practices
- DynamoDB Accelerator
- Amazon DynamoDB best practices
- Amazon Redshift Spectrum best practices
- Prestazioni di Amazon RedShift
- Database cloud con AWS
- Amazon RDS Performance Insights

Video correlati:

- AWS re:Invent 2022 - Monitoraggio delle prestazioni con Amazon e RDS Aurora, con Autodesk
- Monitoraggio e ottimizzazione delle prestazioni del database con Amazon DevOps Guru per Amazon RDS
- AWS re:Invent 2023 - Cosa c'è di nuovo con lo storage di file AWS
- AWS re:Invent 2023 - Scopri di più su Amazon DynamoDB
- AWS re:Invent 2023 - Creazione e ottimizzazione di un data lake su Amazon S3
- AWS re:Invent 2023 - Cosa c'è di nuovo con l'archiviazione dei file AWS
- AWS re:Invent 2023 - Scopri di più su Amazon DynamoDB
- Best practice per il monitoraggio dei carichi di lavoro Redis su Amazon ElastiCache

Esempi correlati:

- Framework di raccolta dei parametri di ingestione del set di dati AWS
- Workshop sul RDS monitoraggio di Amazon
- AWS Workshop sui database appositamente progettato

PERF03-BP04 Implementazione di strategie per migliorare le prestazioni delle query nell'archivio dati

Implementa le strategie per ottimizzare i dati e migliorare le query sui dati in modo da consentire una maggiore scalabilità e prestazioni più efficienti per il tuo carico di lavoro.

Anti-pattern comuni:

- Non suddividi i dati in partizioni nel tuo datastore.
- I dati vengono archiviati in un solo formato di file nel tuo datastore.
- Non usi gli indici nel tuo datastore.

Vantaggi dell'adozione di questa best practice: l'ottimizzazione delle prestazioni dei dati e delle query si traduce in maggiore efficienza, costi inferiori e migliore esperienza utente.

Livello di rischio associato se questa best practice non fosse adottata: medio

Guida all'implementazione

L'ottimizzazione di dati e query è un aspetto critico dell'efficienza delle prestazioni in un datastore, poiché influisce sulle prestazioni e sulla reattività dell'intero carico di lavoro cloud. Le query non ottimizzate possono comportare un maggiore utilizzo delle risorse e rallentamenti, riducendo così l'efficienza complessiva di un datastore.

L'ottimizzazione dei dati include diverse tecniche per garantire prestazioni efficienti per l'archiviazione di dati e il relativo accesso. Ciò aiuta anche a migliorare le prestazioni delle query in un datastore. Le strategie chiave includono il partizionamento, la compressione e la denormalizzazione dei dati, che contribuiscono a ottimizzare i dati sia per l'archiviazione che per l'accesso.

Passaggi dell'implementazione

- Esamina e analizza le query sui dati critiche che vengono eseguite nel tuo datastore.
- Individua le query lente del tuo datastore e utilizza i piani di query per comprenderne lo stato attuale.

  - Analisi del piano di query in Amazon Redshift
  - Usare EXPLAIN e EXPLAIN ANALYZE in Athena
- Implementa le strategie per migliorare le prestazioni delle query. Ecco alcune strategie chiave:
  - Utilizzando un formato di file colonnare (come Parquet o). ORC
  - Compressione dei dati nel datastore per ridurre lo spazio di archiviazione e il funzionamento di I/O.
  - Partizionamento dei dati per suddividere i dati in parti più piccole e ridurre i tempi di analisi dei dati.
    - Partizionamento dei dati in Athena
    - Partitions and data distribution
  - Indicizzazione dei dati sulle colonne comuni della query.
  - Uso delle viste materializzate per le domande frequenti.
    - Understanding materialized views
    - Creating materialized views in Amazon Redshift
  - Scelta dell'operazione di unione corretta per la query. Quando unisci due tabelle, specifica la tabella più grande sul lato sinistro dell'unione e la tabella più piccola sul lato destro.
  - Miglioramento della latenza e riduzione del numero di operazioni di I/O del database grazie alla soluzione di cache distribuita.
  - Manutenzione regolare, ad esempio vacuum, reindicizzazione ed esecuzione di statistiche.
- La sperimentazione e i test delle strategie in un ambiente non di produzione.

Risorse

Documenti correlati:

- Best practice di Amazon Aurora
- Prestazioni di Amazon RedShift
- Amazon Athena top 10 performance tips
- AWS Database Caching
- Best practice per l'implementazione di Amazon ElastiCache
- Partizionamento dei dati in Athena

Video correlati:

- AWS re:Invent 2023: best practice per l'ottimizzazione dei costi AWS di storage
- AWS re:Invent 2022 - Monitoraggio delle prestazioni con Amazon e RDS Aurora, con Autodesk
- Optimize Amazon Athena Queries with New Query Analysis Tools

Esempi correlati:

- Amazon S3 Select - Querying data without servers or databases
- AWS Workshop sui database appositamente progettati

PERF03-BP05 Implementare modelli di accesso ai dati che utilizzano la memorizzazione nella cache

Implementa modelli di accesso che possano trarre vantaggio dalla memorizzazione dei dati nella cache per il recupero rapido dei dati a cui si accede di frequente.

Anti-pattern comuni:

- Memorizzare nella cache dati che cambiano in maniera frequente.
- Fare affidamento sui dati memorizzati nella cache come se fossero archiviati in modo duraturo e sempre disponibili.
- Non tenere conto della coerenza dei dati memorizzati nella cache.
- Non monitorare l'efficienza dell'implementazione della cache.

Vantaggi dell'adozione di questa best practice: l'archiviazione dei dati in una cache può migliorare la latenza di lettura, il throughput, l'esperienza utente e l'efficienza complessiva, oltre a ridurre i costi.

Livello di rischio associato se questa best practice non fosse adottata: medio

Guida all'implementazione

Una cache è un componente software o hardware progettato per archiviare dati in modo che le richieste future degli stessi dati possano essere soddisfatte più velocemente o in modo più efficiente. I dati memorizzati in una cache possono essere ricostruiti in caso di perdita, ripetendo un calcolo precedente o recuperandolo da un altro datastore.

La memorizzazione dei dati nella cache può essere una delle strategie più efficaci per migliorare le prestazioni complessive delle applicazioni e ridurre il carico sulle origini dati primarie sottostanti.

È possibile memorizzare i dati nella cache a più livelli dell'applicazione, ad esempio all'interno dell'applicazione che effettua chiamate remote, operazione nota come memorizzazione nella cache lato client, o mediante un servizio secondario veloce per l'archiviazione dei dati, operazione nota come memorizzazione nella cache remota.

Memorizzazione nella cache lato client

Con la memorizzazione nella cache lato client, ogni client (un'applicazione o un servizio che interroga il datastore di backend) può archiviare localmente i risultati delle proprie query uniche per un periodo di tempo specificato. Ciò può ridurre il numero di richieste a un datastore attraverso la rete perché viene controllata prima la cache del client locale. Se questa non contiene risultati, l'applicazione può interrogare il datastore e archiviare tali risultati localmente. Questo modello consente a ciascun client di archiviare i dati nella sede più vicina possibile (il client stesso), garantendo così la latenza più bassa possibile. I client possono inoltre continuare a eseguire query quando il datastore di backend non è disponibile, aumentando la disponibilità dell'intero sistema.

Uno svantaggio di questo approccio è che quando sono coinvolti più client, potrebbero archiviare localmente gli stessi dati memorizzati nella cache. Ciò si traduce in un utilizzo duplicato dell'archiviazione e nell'incoerenza dei dati tra questi client. Può accadere che un client memorizzi nella cache i risultati di una query e un minuto dopo un altro client esegua la stessa query ottenendo un risultato diverso.

Memorizzazione nella cache remota

Come soluzione al problema della duplicazione dei dati tra client, è possibile utilizzare un servizio esterno veloce o la memorizzazione nella cache remota per archiviare i dati sottoposti a query. Anziché controllare un datastore locale, ogni client controllerà la cache remota prima di interrogare il datastore di backend. Questa strategia consente di ottenere risposte più coerenti tra i client, una migliore efficienza dei dati archiviati e un volume maggiore di dati memorizzati nella cache, perché lo spazio di archiviazione si dimensiona in maniera indipendente dai client.

Lo svantaggio di una cache remota è che l'intero sistema può registrare una latenza più elevata, poiché è necessario un hop di rete aggiuntivo per controllare la cache remota. Per migliorare la latenza, è possibile utilizzare la memorizzazione nella cache lato client insieme alla memorizzazione nella cache remota, eseguendo così una memorizzazione nella cache su più livelli.

Passaggi dell'implementazione

- Identifica i database APIs e i servizi di rete che potrebbero trarre vantaggio dalla memorizzazione nella cache. I servizi che hanno carichi di lavoro di lettura elevati, hanno un read-to-write rapporto elevato o sono costosi da scalare sono candidati alla memorizzazione nella cache.
  - Database Caching
  - Abilitare la API memorizzazione nella cache per migliorare la reattività
- Identifica il tipo di strategia di memorizzazione nella cache più adatto al tuo modello di accesso.
  - Caching strategies
  - AWS Caching Solutions
- Attieniti alle best practice sulla memorizzazione nella cache per il tuo archivio dati.
- Configura una strategia di invalidazione della cache, ad esempio a time-to-live (TTL), per tutti i dati che bilanci l'aggiornamento dei dati e riduca la pressione sul datastore di backend.
- Abilita funzionalità quali tentativi di connessione automatici, backoff esponenziale, timeout lato client e pool di connessioni nel client, se disponibili, che possono migliorare prestazioni e affidabilità.
  - Migliori pratiche: client Redis e Amazon ElastiCache (OSSRedis)
- Monitora la percentuale di riscontri nella cache con un obiettivo dell'80% o superiore. Valori inferiori possono indicare una dimensione della cache insufficiente o un modello di accesso che non sfrutta la memorizzazione nella cache.
  - Which metrics should I monitor?
  - Le migliori pratiche per il monitoraggio dei carichi di lavoro Redis su Amazon ElastiCache
  - Monitoraggio delle best practice con Amazon ElastiCache (RedisOSS) tramite Amazon CloudWatch
- Implementa la replica dei dati per eliminare il carico delle letture per più istanze e migliorare prestazioni e disponibilità della lettura dei dati.

Risorse

Documenti correlati:

- Utilizzo dell'obiettivo Amazon ElastiCache Well-Architected
- Monitoraggio delle best practice con Amazon ElastiCache (RedisOSS) tramite Amazon CloudWatch

- Quali parametri è opportuno monitorare?
- Prestazioni su larga scala con il ElastiCache white paper di Amazon
- Sfide e strategie del caching

Video correlati:

- Percorso di ElastiCache apprendimento su Amazon
- Progetta per il successo con le ElastiCache best practice di Amazon
- AWS re:Invent 2020 - Progetta per il successo con le best practice di Amazon ElastiCache
- AWS re:Invent 2023 - [] LAUNCH Presentazione di Amazon Serverless ElastiCache
- AWS re:Invent 2022 - 5 ottimi modi per reimmaginare il tuo livello di dati con Redis
- AWS re:Invent 2021 - Approfondimento su Amazon ElastiCache (Redis) OSS

Esempi correlati:

- Potenziamento delle prestazioni SQL del mio database con Amazon ElastiCache (RedisOSS)

Reti e distribuzione di contenuti

Questions

- PERF4. In che modo selezioni e configuri le risorse di rete nel carico di lavoro?

PERF4. In che modo selezioni e configuri le risorse di rete nel carico di lavoro?

La soluzione di rete ottimale per un carico di lavoro varia in base a latenza, requisiti di throughput, jitter e larghezza di banda. I vincoli fisici, ad esempio le risorse utente o on-premises, determinano le opzioni di posizione. Questi vincoli possono essere compensati con le posizioni edge o la collocazione delle risorse.

Best practice

- PERF04-BP01 Scopri come la rete influisce sulle prestazioni
- PERF04-BP02 Valuta le funzionalità di rete disponibili
- PERF04-BP03 Scegli la connettività dedicata appropriata o per il tuo carico di lavoro VPN
- PERF04-BP04 Usa il bilanciamento del carico per distribuire il traffico su più risorse

- PERF04-BP05 Scegli i protocolli di rete per migliorare le prestazioni
- PERF04-BP06 Scegli la posizione del carico di lavoro in base ai requisiti di rete
- PERF04-BP07 Ottimizza la configurazione di rete in base a metriche

PERF04-BP01 Scopri come la rete influisce sulle prestazioni

Analizza e comprendi in che modo le decisioni correlate alla rete influiscono sul carico di lavoro per fornire prestazioni efficienti e una migliore esperienza utente.

Anti-pattern comuni:

- Tutto il traffico passa attraverso i data center esistenti.
- Si instrada tutto il traffico attraverso i firewall centrali anziché utilizzare strumenti di sicurezza di rete nativi del cloud.
- Effettua il provisioning delle AWS Direct Connect connessioni senza comprendere i requisiti di utilizzo effettivi.
- Quando si definiscono le soluzioni di rete, non si considerano le caratteristiche del carico di lavoro e l'overhead della crittografia.
- Per le soluzioni di rete nel cloud si utilizzano concetti e strategie on-premises.

Vantaggi dell'adozione di questa best practice: la comprensione dell'impatto della rete sulle prestazioni del carico di lavoro ti aiuta a identificare i potenziali colli di bottiglia, migliorare l'esperienza dell'utente, aumentare l'affidabilità e ridurre la manutenzione operativa al variare del carico di lavoro.

Livello di rischio associato se questa best practice non fosse adottata: elevato

Guida all'implementazione

La rete è responsabile della connettività tra componenti dell'applicazione, servizi cloud, reti edge e dati on-premises e quindi può avere un forte impatto sulle prestazioni dei carichi di lavoro. Oltre alle prestazioni del carico di lavoro, l'esperienza dell'utente può essere influenzata anche da latenza della rete, larghezza di banda, protocolli, posizione, congestione della rete, jitter, throughput e regole di instradamento.

Predisponi un elenco documentato dei requisiti di rete del carico di lavoro, tra cui latenza, dimensione dei pacchetti, regole di instradamento, protocolli e modelli di traffico di supporto. Esamina le soluzioni di rete disponibili e individua il servizio che soddisfi le caratteristiche di rete del proprio carico

di lavoro. Le reti basate sul cloud possono essere ricostruite rapidamente, quindi l'evoluzione dell'architettura di rete nel tempo è necessaria per migliorare l'efficienza delle prestazioni.

Passaggi dell'implementazione:

- Definisci e documenta i requisiti di prestazioni di rete, tra cui metriche come latenza di rete, larghezza di banda, protocolli, posizioni, modelli di traffico (picchi e frequenza), throughput, crittografia, ispezione e regole di instradamento.
- Scopri i principali servizi AWS di rete come VPCsElastic Load Balancing (ELB) e Amazon Route 53. AWS Direct Connect
- Acquisisci le seguenti caratteristiche di rete fondamentali:

Caratteristiche	Strumenti e metriche
Caratteristiche fondamentali della rete	• VPCRegistri di flusso • AWS Transit Gateway Log di flusso • AWS Transit Gateway metriche • AWS PrivateLink metriche
Caratteristiche di rete dell'applicazione	• Elastic Fabric Adapter (EFA) • AWS App Mesh metriche • Metriche di Amazon API Gateway
Caratteristiche della rete edge	• CloudFront Metriche Amazon • Parametri di Amazon Route 53 • AWS Global Accelerator metriche
Caratteristiche della rete ibrida	• AWS Direct Connect metriche • AWS Site-to-Site VPN metriche • AWS Client VPN metriche • Cloud AWS WANmetriche
Caratteristiche della sicurezza di rete	• AWS Shield e AWS WAF metriche AWS Network Firewall
Caratteristiche del tracciamento	• AWS X-Ray

Caratteristiche	Strumenti e metriche
	• VPCReachability Analyzer • Strumento di analisi degli accessi alla rete • Amazon Inspector • Amazon CloudWatch RUM

- Esegui il benchmark e testa le prestazioni della rete:
  - Effettua un benchmark del throughput di rete, poiché alcuni fattori possono influire sulle prestazioni EC2 della rete Amazon quando le istanze sono uguali. VPC Misura la larghezza di banda di rete tra le istanze di Amazon EC2 Linux nella stessa istanza. VPC
  - Effettua test di carico per sperimentare soluzioni e opzioni di rete.

Risorse

Documenti correlati:

- Application Load Balancer
- EC2Rete avanzata su Linux
- EC2Rete avanzata su Windows
- EC2Gruppi di collocamento
- Abilitazione di reti avanzate con Elastic Network Adapter (ENA) su istanze Linux
- Network Load Balancer
- Prodotti di rete con AWS
- Gateway di transito
- Transitioning to latency-based routing in Amazon Route 53
- VPCEndpoint

Video correlati:

- AWS re:Invent 2023 - fondamenti per il networking AWS
- AWS re:Invent 2023 - Cosa può fare il networking per la tua applicazione?
- AWS re:Invent 2023 - Design avanzati e nuove funzionalità VPC
- AWS re:Invent 2023 - Una guida per sviluppatori al cloud networking

- AWS re:Invent 2019 - Connettività e architetture di rete ibride AWSAWS
- AWS re:Invent 2019 - Ottimizzazione delle prestazioni di rete per le istanze Amazon EC2
- AWS Summit Online - Migliora le prestazioni della rete globale per le applicazioni
- AWS re:Invent 2020 - Migliori pratiche e suggerimenti per il networking con il Well-Architected Framework
- AWS re:Invent 2020 - migliori pratiche di rete nelle migrazioni su larga scala AWS

Esempi correlati:

- AWS Transit Gateway e soluzioni di sicurezza scalabili
- AWS Workshop di networking
- Hands-on Network Firewall Workshop
- Osservazione e diagnosi della rete su AWS
- Individuazione e risoluzione degli errori di configurazione della rete su AWS

PERF04-BP02 Valuta le funzionalità di rete disponibili

Valuta le funzionalità di rete nel cloud che possono aumentare le prestazioni. Misura l'impatto di tali funzionalità attraverso test, parametri e analisi. Ad esempio, sfrutta le funzionalità a livello di rete disponibili per ridurre latenza, distanza di rete o jitter.

Anti-pattern comuni:

- Rimani all'interno di una regione perché è lì che si trova fisicamente la tua sede centrale.
- Utilizzi i firewall anziché i gruppi di sicurezza per filtrare il traffico.
- Fai una pausa TLS per ispezionare il traffico anziché affidarti a gruppi di sicurezza, policy degli endpoint e altre funzionalità native del cloud.
- Utilizzi solo la segmentazione basata su sottoreti anziché i gruppi di sicurezza.

Vantaggi dell'adozione di questa best practice la valutazione di tutte le funzionalità e le opzioni del servizio consente di ridurre il costo dell'infrastruttura e l'impegno necessario per mantenere il carico di lavoro e aumentare l'assetto di sicurezza generale. Puoi utilizzare il AWS backbone globale per fornire un'esperienza di rete ottimale ai tuoi clienti.

Livello di rischio associato se questa best practice non fosse adottata: elevato

Guida all'implementazione

AWS offre servizi come AWS Global AcceleratorAmazon CloudFront che possono aiutare a migliorare le prestazioni di rete, mentre la maggior parte dei AWS servizi dispone di caratteristiche di prodotto (come la funzionalità Amazon S3 Transfer Acceleration) per ottimizzare il traffico di rete.

Analizza quali opzioni di configurazione relative alla rete sono disponibili e come possono influire sul tuo carico di lavoro. L'ottimizzazione delle prestazioni dipende dalla comprensione del modo in cui queste opzioni interagiscono con l'architettura e dall'impatto che hanno sulle prestazioni misurate e sull'esperienza utente.

Passaggi dell'implementazione

- Crea l'elenco dei componenti del carico di lavoro.
  - Prendi in considerazione l'idea Cloud AWS WANdi utilizzarla per creare, gestire e monitorare la rete della tua organizzazione quando crei una rete globale unificata.
  - Monitora le tue reti globali e principali con i parametri di Amazon CloudWatch Logs. Sfrutta Amazon CloudWatch RUM, che fornisce informazioni utili per identificare, comprendere e migliorare l'esperienza digitale degli utenti.
  - Visualizza la latenza di rete aggregata tra Regioni AWS e le zone di disponibilità, nonché all'interno di ciascuna zona di disponibilità, utilizzala AWS Network Managerper ottenere informazioni dettagliate su come le prestazioni delle applicazioni si relazionano con le prestazioni della rete sottostante. AWS
  - Utilizzate uno strumento esistente per il database di gestione della configurazione (CMDB) o un servizio, AWS Configad esempio, per creare un inventario del carico di lavoro e della sua configurazione.
- Se si tratta di un carico di lavoro esistente, individua e documenta l'analisi di benchmark per le metriche relative alle prestazioni, concentrandoti sui colli di bottiglia e sulle aree da migliorare. Le metriche relative alla rete a livello di prestazioni varieranno a seconda dei requisiti aziendali e delle caratteristiche del carico di lavoro. Come punto di partenza, le seguenti metriche possono essere importanti per la revisione del carico di lavoro: larghezza di banda, latenza, perdita di pacchetti, jitter e ritrasmissioni.
- Se si tratta di un nuovo carico di lavoro, esegui test di carico per individuare i colli di bottiglia delle prestazioni.
- Per tutti i colli di bottiglia di questo tipo individuati, esamina le opzioni di configurazione per le soluzioni in uso per individuare le opportunità di miglioramento delle prestazioni. Consulta le seguenti opzioni e funzionalità di rete fondamentali:

Opportunità di miglioramento	Soluzione
Percorso o instradamenti di rete	Usa lo Strumento di analisi degli accessi alla rete per identificare percorsi o percorsi.
Protocolli di rete	Consulta la sezione PERF04-BP05 Scegli i protocolli di rete per migliorare le prestazioni
Topologia di rete	Valuta i compromessi operativi e prestazio nali tra VPCPeering e AWS Transit Gatewayquando si connettono più account. AWS Transit Gateway semplifica il modo in cui interconnetti tutti i tuoi dispositiviVPCs, che possono estendersi su migliaia di reti locali. Account AWS Condividi il tuo AWS Transit Gateway tra più account utilizzando. AWS Resource Access Manager Consulta la sezione PERF04-BP03 Scegli la connettività dedicata appropriata o per il tuo carico di lavoro VPN

Opportunità di miglioramento	Soluzione
Servizi di rete	AWS Global Acceleratorè un servizio di rete che migliora le prestazioni del traffico degli utenti fino al 60% utilizzando l'infrastruttura di rete AWS globale. Amazon CloudFront può migliorare le prestazioni della distribuzione e della latenza dei contenuti del carico di lavoro a livello globale. Usa Lambda @edge per eseguire funzioni che personalizzano i contenuti per renderli più vicini agli utenti, ridurre la latenza e migliorare le CloudFront prestazioni. Amazon Route 53 offre opzioni di instradamento basato sulla latenza, instradamento basato sulla geolocalizzazione, instradamento basato sulla geoprossimità e instradamento basato su IP per migliorare le prestazioni del tuo carico di lavoro per un pubblico globale. Rivedi il traffico del carico di lavoro e la posizione dell'utente quando il carico di lavoro è distribuito a livello globale per individuare quale opzione di instradamento è in grado di ottimizzare le prestazioni del carico di lavoro.

Opportunità di miglioramento	Soluzione
Funzionalità delle risorse di archiviazione	Amazon S3 Transfer Acceleration è una funzionalità che consente agli utenti esterni di beneficiare CloudFront delle ottimizzazioni di rete per caricare dati su Amazon S3. Ciò migliora le caratteristiche di trasferimento di grandi quantità di dati da posizioni remote prive di connettività dedicata al Cloud AWS. I punti di accesso multi-regione di Amazon S3 rappresentano una funzionalità che replica i contenuti in più regioni e semplifica il carico di lavoro fornendo un punto di accesso. Quando viene utilizzato un punto di accesso multi-reg ione, puoi richiedere o scrivere dati in Amazon S3 con il servizio che identifica il bucket con latenza più bassa.

Opportunità di miglioramento	Soluzione
Funzionalità delle risorse di calcolo	Le interfacce di rete elastiche (ENA) utilizzate dalle EC2 istanze, dai contenitori e dalle funzioni Lambda di Amazon sono limitate in base al flusso. Controlla i tuoi gruppi di collocamento per ottimizzare la velocità di trasmissione della rete. EC2 Per evitare colli di bottiglia a livello di flusso, progetta l'applicazione in modo che utilizzi più flussi. Per monitorare e ottenere visibilità sulle metriche di rete relative all'elaborazione, usa CloudWatch Metrics ed ethtool. Il `ethtool` comando è incluso nel ENA driver ed espone metriche aggiuntive relative alla rete che possono essere pubblicate come metriche personalizzate su. CloudWatch Amazon Elastic Network Adapters (ENA) forniscono un'ulteriore ottimizzazione offrendo un throughput migliore per le istanze all'interno di un gruppo di collocamento di cluster. Elastic Fabric Adapter (EFA) è un'interfaccia di rete per EC2 istanze Amazon che consente di eseguire carichi di lavoro che richiedono alti livelli di comunicazioni tra nodi su larga scala. AWS Le istanze EBS ottimizzate per Amazon utilizzano uno stack di configurazione ottimizzato e forniscono capacità aggiuntiva dedicata per aumentare l'I/O di Amazon. EBS

Risorse

Documenti correlati:

- Application Load Balancer
- EC2Rete avanzata su Linux
- EC2Rete avanzata su Windows
- EC2Gruppi di collocamento
- Abilitazione di reti avanzate con Elastic Network Adapter (ENA) su istanze Linux
- Network Load Balancer
- Prodotti di rete con AWS
- Transitioning to Latency-Based Routing in Amazon Route 53
- VPCEndpoint
- Log di flusso VPC

Video correlati:

- AWS re:Invent 2023 — Pronti per il futuro? Designing networks for growth and flexibility
- AWS re:Invent 2023 — Design avanzati e nuove funzionalità VPC
- AWS re:Invent 2023: guida per sviluppatori al cloud networking
- AWS re:Invent 2022 — Approfondisci l'infrastruttura di rete AWS
- AWS re:Invent 2019 — Connettività e architetture di rete ibride AWSAWS
- AWS re:Invent 2018 — Ottimizzazione delle prestazioni di rete per le istanze Amazon EC2
- AWS Global Accelerator

Esempi correlati:

- AWS Transit Gateway e soluzioni di sicurezza scalabili
- AWS Workshop di networking
- Observing and diagnosing your network
- Individuazione e risoluzione degli errori di configurazione della rete su AWS

PERF04-BP03 Scegli la connettività dedicata appropriata o per il tuo carico di lavoro VPN

Quando hai bisogno di una connettività ibrida per connettere risorse on-premises e cloud, assicurati di avere una larghezza di banda adeguata per soddisfare i tuoi requisiti di prestazione. Fai una

stima dei requisiti di larghezza di banda e latenza per il carico di lavoro ibrido. I valori calcolati determineranno le tue esigenze di dimensionamento.

Anti-pattern comuni:

- Valutate solo VPN soluzioni per i vostri requisiti di crittografia di rete.
- Non vengono valutate opzioni di backup o di connettività ridondante.
- Non è possibile identificare tutti i requisiti del carico di lavoro (esigenze di crittografia, protocollo, larghezza di banda e traffico).

Vantaggi dell'adozione di questa best practice: la selezione e la configurazione di soluzioni di connettività appropriate migliorano l'affidabilità del carico di lavoro e massimizzano le prestazioni. Identificando i requisiti del carico di lavoro, pianificando in anticipo e valutando le soluzioni ibride, è possibile ridurre al minimo le costose modifiche alla rete fisica e il sovraccarico operativo, aumentando al contempo i costi. time-to-value

Livello di rischio associato se questa best practice non fosse adottata: elevato

Guida all'implementazione

Sviluppa un'architettura di rete ibrida basata sui tuoi requisiti di larghezza di banda. AWS Direct Connect consente di connettere la rete on-premises in privato con AWS. È utile quando hai bisogno di larghezza di banda elevata, bassa latenza e di mantenere le prestazioni coerenti. Una VPN connessione stabilisce una connessione sicura su Internet. Viene utilizzata quando è necessaria solo una connessione temporanea, quando il costo è un fattore importante o come misura di contingenza in attesa che venga stabilita una connettività di rete fisica resiliente mentre AWS Direct Connectè in uso.

Se i requisiti di larghezza di banda sono elevati, potresti prendere in considerazione più servizi AWS Direct Connect o VPN più servizi. È possibile bilanciare il carico del traffico tra i servizi, anche se non consigliamo il bilanciamento del carico tra AWS Direct Connect e a VPN causa delle differenze di latenza e larghezza di banda.

Passaggi dell'implementazione

- Calcola i requisiti di larghezza di banda e latenza delle tue app esistenti.
  - Per i carichi di lavoro esistenti che stanno per essere trasferiti AWS, sfrutta i dati dei tuoi sistemi di monitoraggio della rete interna.

  - Per i carichi di lavoro nuovi o esistenti per i quali non sono disponibili dati di monitoraggio, consulta i proprietari dei prodotti per definire metriche sulle prestazioni adeguate e offrire un'esperienza utente soddisfacente.
- Seleziona una connessione dedicata o VPN come opzione di connettività. In base a tutti i requisiti del carico di lavoro (crittografia, larghezza di banda e traffico), puoi scegliere AWS Direct Connect o AWS VPN(o entrambi). Il diagramma seguente può aiutarti a scegliere il tipo di connessione appropriato.
  - AWS Direct Connect fornisce connettività dedicata all'ambiente AWS da 50 Mbps fino a 100 Gbps, utilizzando connessioni dedicate od ospitate. In questo modo, disporrai di latenza gestita e controllata, nonché di larghezza di banda assegnata, in modo che il carico di lavoro possa connettersi con efficienza ad altri ambienti. Utilizzando AWS Direct Connect i partner, è possibile disporre di end-to-end connettività da più ambienti, fornendo una rete estesa con prestazioni costanti. AWS offre la scalabilità della larghezza di banda della connessione Direct Connect utilizzando 100 Gbps nativi, link aggregation group (LAG) o BGP equal-cost multipath (). ECMP
  - AWS Site-to-Site VPNFornisce un servizio gestito VPN che supporta la sicurezza del protocollo Internet (). IPsec Quando viene creata una VPN connessione, ogni VPN connessione include due tunnel per un'elevata disponibilità.
- Segui AWS la documentazione per scegliere l'opzione di connettività appropriata:
  - Se decidi di utilizzarla AWS Direct Connect, seleziona la larghezza di banda appropriata per la tua connettività.
  - Se utilizzi una connessione AWS Site-to-Site VPN tra più postazioni per connetterti a una Regione AWS, utilizza una Site-to-SiteVPNconnessione accelerata per avere l'opportunità di migliorare le prestazioni della rete.
  - Se la progettazione della rete prevede una IPSec VPN connessione via cavo AWS Direct Connect, prendete in considerazione l'utilizzo di Private IP VPN per migliorare la sicurezza e ottenere la segmentazione.AWS Site-to-Site VPNL'IP privato viene distribuito sopra l'interfaccia virtuale di transito ()VIF.
  - AWS Direct Connect SiteLinkconsente di creare connessioni ridondanti e a bassa latenza tra i data center in tutto il mondo inviando i dati lungo il percorso più veloce tra le sedi, evitando così di farlo.AWS Direct Connect Regioni AWS
- Convalida la configurazione della connettività prima di eseguire l'implementazione in produzione. Esegui test di sicurezza e prestazioni per assicurarti di soddisfare i requisiti di larghezza di banda, affidabilità, latenza e conformità.
- Monitora regolarmente le prestazioni e l'utilizzo della connettività e ottimizzali, se necessario.

Diagramma di flusso per le prestazioni deterministiche

Risorse

Documenti correlati:

- Prodotti di rete con AWS
- AWS Transit Gateway
- VPCEndpoint
- Creazione di un'infrastruttura multirete scalabile e sicura VPC AWS
- Cliente VPN

Video correlati:

- AWS re:Invent 2023 — Creazione di una connettività di rete ibrida con AWS
- AWS re:Invent 2023 — Connettività remota sicura a AWS
- AWS re:Invent 2022 — Ottimizzazione delle prestazioni con Amazon CloudFront
- AWS re:Invent 2019 — Connettività e architetture di rete ibride AWSAWS
- AWS re:Invent 2020 — Connect AWS Transit Gateway

Esempi correlati:

- AWS Transit Gateway e soluzioni di sicurezza scalabili
- AWS Workshop di networking

PERF04-BP04 Usa il bilanciamento del carico per distribuire il traffico su più risorse

Distribuisci il traffico tra varie risorse o servizi affinché il carico di lavoro possa trarre vantaggio dall'elasticità fornita dal cloud. Puoi anche utilizzare il bilanciamento del carico per la terminazione dell'offloading della crittografia al fine di migliorare le prestazioni, l'affidabilità e gestire e instradare il traffico in modo efficiente.

Anti-pattern comuni:

- Scelta del tipo di sistema di bilanciatore del carico senza tenere conto dei requisiti del carico di lavoro.
- Mancato utilizzo delle funzionalità del bilanciatore del carico per l'ottimizzazione delle prestazioni.
- Esposizione diretta del carico di lavoro a Internet senza un bilanciatore del carico.
- Instradi tutto il traffico Internet attraverso i bilanciatori del carico esistenti.
- Utilizzi il bilanciamento del TCP carico generico e fai in modo che ogni nodo di calcolo gestisca la crittografia. SSL

Vantaggi dell'adozione di questa best practice: un bilanciatore del carico gestisce il carico variabile del traffico dell'applicazione in una o più zone di disponibilità e consente alta disponibilità, dimensionamento automatico e un migliore utilizzo del carico di lavoro.

Livello di rischio associato se questa best practice non fosse adottata: elevato

Guida all'implementazione

I bilanciatori del carico operano come punto di ingresso per il carico di lavoro, dal quale distribuiscono il traffico alle destinazioni di backend, come istanze di calcolo o container per migliorarne l'utilizzo.

La scelta del tipo corretto di bilanciatore del carico è il primo passaggio per ottimizzare l'architettura. Inizia elencando le caratteristiche del carico di lavoro, ad esempio il protocollo (ad esempio TCPHTTP,TLS, o WebSockets), il tipo di destinazione (come istanze, contenitori o serverless), i requisiti dell'applicazione (come connessioni a lunga durata, autenticazione utente o persistenza) e il posizionamento (come regione, zona locale, Outpost o isolamento zonale).

AWS fornisce diversi modelli per le applicazioni per utilizzare il bilanciamento del carico. Application Load Balancer è la soluzione ideale per il bilanciamento del carico HTTP e del HTTPS traffico e fornisce un routing avanzato delle richieste mirato alla distribuzione di architetture applicative moderne, inclusi microservizi e contenitori.

Network Load Balancer è la soluzione ideale per il bilanciamento del carico del TCP traffico laddove sono richieste prestazioni estreme. È in grado di gestire milioni di richieste al secondo, mantenendo al contempo latenze ridottissime. Inoltre, è ottimizzato per la gestione degli schemi di traffico improvvisi e incostanti.

Elastic Load Balancing offre la gestione e SSL la TLS decrittografia integrate dei certificati, che ti offrono la flessibilità necessaria per gestire centralmente SSL le impostazioni del sistema di bilanciamento del carico e ridurre il carico di lavoro CPU intensivo.

Dopo aver scelto il bilanciatore del carico appropriato, puoi iniziare a utilizzarne le funzionalità per ridurre la quantità di attività che deve svolgere il backend per distribuire il traffico.

Ad esempio, utilizzando sia Application Load Balancer (ALB) che Network Load Balancer NLB (), è possibile SSL eseguireTLS/encryption offloading, un'opportunità per evitare che CPU l'handshake così TLS impegnativo venga completato dai target e anche per migliorare la gestione dei certificati.

Quando configuriSSL/TLSoffloading nel tuo sistema di bilanciamento del carico, quest'ultimo diventa responsabile della crittografia del traffico da e verso i client, mentre consegna il traffico non crittografato ai tuoi backend, liberando le tue risorse di backend e migliorando i tempi di risposta per i client.

Application Load Balancer può anche servire il traffico HTTP /2 senza bisogno di supportarlo sui tuoi obiettivi. Questa semplice decisione può migliorare i tempi di risposta dell'applicazione, poiché HTTP /2 utilizza TCP le connessioni in modo più efficiente.

Nel definire l'architettura, è bene tenere conto dei requisiti di latenza del carico di lavoro. Ad esempio, se un'applicazione è sensibile alla latenza, è possibile scegliere di usare Network Load Balancer, che offre latenze estremamente ridotte. In alternativa, è possibile decidere di avvicinare il carico di lavoro ai clienti sfruttando Application Load Balancer nelle zone locali AWS o addirittura AWS Outposts.

Un altro aspetto di cui tenere conto per i carichi di lavoro sensibili alla latenza è il bilanciamento del carico tra zone. Con il bilanciamento del carico tra zone, ogni nodo del bilanciatore del carico distribuisce il traffico tra le destinazioni registrate in tutte le zone di disponibilità autorizzate.

Usa Auto Scaling integrato con il bilanciatore del carico. Uno degli aspetti principali di un sistema con prestazioni efficienti riguarda il dimensionamento corretto delle risorse backend. A questo scopo, puoi utilizzare integrazioni dei bilanciatori del carico per le risorse di destinazione backend. Usando l'integrazione dei bilanciatori del carico con gruppi Auto Scaling, le destinazioni vengono aggiunte o rimosse nel e dal bilanciatore del carico in base alle esigenze, in risposta al traffico in ingresso. I sistemi di bilanciamento del carico possono anche integrarsi con Amazon e ECSAmazon EKS per carichi di lavoro containerizzati.

- Amazon ECS - Bilanciamento del carico di servizio
- Bilanciamento del carico delle applicazioni su Amazon EKS
- Bilanciamento del carico di rete su Amazon EKS

Passaggi dell'implementazione

- Definisci i tuoi requisiti di bilanciamento del carico, tra cui volume di traffico, disponibilità e scalabilità delle applicazioni.
- Scegli il tipo di sistema di bilanciatore del carico del carico giusto per la tua applicazione.
  - Usa Application Load Balancer per i carichi di lavoro HTTP HTTPS /.
  - Usa Network Load Balancer per HTTP carichi non di lavoro eseguiti su o. TCP UDP
  - Utilizza una combinazione di entrambi (ALBcome obiettivo di NLB) se desideri sfruttare le funzionalità di entrambi i prodotti. Ad esempio, puoi farlo se desideri utilizzare lo statico IPs di NLB insieme al routing basato sull'HTTPintestazione da ALB o se desideri esporre il tuo HTTP carico di lavoro a un. AWS PrivateLink
  - Per un confronto completo dei sistemi di bilanciamento del carico, consulta la sezione Confronto dei prodotti. ELB
- Se possibile, usaSSL/TLSoffloading.
  - ConfiguraHTTPS/TLSlistener con Application Load Balancer e Network Load Balancer integrati con. AWS Certificate Manager
  - Tieni presente che alcuni carichi di lavoro potrebbero richiedere la end-to-end crittografia per motivi di conformità. In questo caso, è necessario consentire la crittografia nelle destinazioni.
  - Per le migliori pratiche di sicurezza, consulta SEC09-BP02 Applica la crittografia in transito.
- Seleziona l'algoritmo di routing corretto (solo). ALB
  - L'algoritmo di instradamento può fare la differenza per quanto riguarda l'uso corretto delle destinazioni backend e, di conseguenza, l'impatto sulle prestazioni. Ad esempio, ALB offre due opzioni per gli algoritmi di routing:
  - Numero minimo di richieste in sospeso: usa questa opzione per ottenere una migliore distribuzione del carico nelle destinazioni backend nei casi in cui le richieste per l'applicazione variano per complessità o le destinazioni variano per capacità di elaborazione.
  - Round robin: usa questa opzione quando le richieste e le destinazioni sono simili o se devi distribuire equamente le richieste tra le destinazioni.
- Valuta se usare l'isolamento tra zone o quello zonale.
  - Disattiva l'isolamento tra zone (usando l'isolamento zonale) per migliorare la latenza e in caso di domini con errori di zona. È disattivato per impostazione predefinita in NLB e ALBpuoi disattivarlo per gruppo target.

    - Attiva l'isolamento tra zone per ottenere disponibilità e flessibilità maggiori. Per impostazione predefinita, l'opzione cross-zone è attivata per ogni gruppo target ALB e NLBpuò essere attivata in base al gruppo target.
- Attiva HTTP keep-alive per i tuoi HTTP carichi di lavoro (solo). ALB Con questa funzionalità, il load balancer può riutilizzare le connessioni di backend fino alla scadenza del timeout keep-alive, migliorando i tempi di HTTP richiesta e risposta e riducendo anche l'utilizzo delle risorse per gli obiettivi di backend. Per informazioni dettagliate su come eseguire questa operazione per Apache e Nginx, vedi Quali sono le impostazioni ottimali per l'utilizzo di Apache o come server di backend? NGINX ELB
- Attiva il monitoraggio del tuo bilanciatore del carico.
    - Attiva i log di accesso per Application Load Balancer e Network Load Balancer.
    - I campi principali da considerare sono, e. ALB `request_processing_time` `request_processing_time` `response_processing_time`
    - I principali campi da prendere in considerazione NLB sono `connection_time` `etls_handshake_time`.
    - Preparati a eseguire query sui log quando necessario. Puoi usare Amazon Athena per interrogare sia i ALBlog che i log. NLB
    - Crea allarmi per metriche relative alle prestazioni, ad esempio per. `TargetResponseTime` ALB

Risorse

Documenti correlati:

- ELBconfronto tra prodotti
- AWS Infrastruttura globale
- Improving Performance and Reducing Cost Using Availability Zone Affinity
- Step by step for Log Analysis with Amazon Athena
- Querying Application Load Balancer logs
- Monitor your Application Load Balancers
- Monitor your Network Load Balancer
- Use Elastic Load Balancing to distribute traffic across the instances in your Auto Scaling group

Video correlati:

- AWS re:Invent 2023: Cosa può fare il networking per la tua applicazione?
- AWS RE:InForce 20: Come utilizzare Elastic Load Balancing per migliorare il livello di sicurezza su larga scala
- AWS re:Invent 2018: Elastic Load Balancing: approfondimenti e best practice
- AWS re:Invent 2021 - Come scegliere il bilanciamento del carico giusto per i tuoi carichi di lavoro AWS
- AWS re:Invent 2019: ottieni il massimo da Elastic Load Balancing per diversi carichi di lavoro

Esempi correlati:

- Gateway Load Balancer
- CDKed AWS CloudFormation esempi per l'analisi dei log con Amazon Athena

PERF04-BP05 Scegli i protocolli di rete per migliorare le prestazioni

Prendi decisioni sui protocolli per la comunicazione tra sistemi e reti in base all'impatto sulle prestazioni del carico di lavoro.

Esiste una relazione tra latenza e larghezza di banda per ottenere il throughput desiderato. Se il trasferimento dei file utilizza Transmission Control Protocol (TCP), latenze più elevate molto probabilmente ridurranno la velocità effettiva complessiva. Esistono approcci per risolvere questo problema con l'TCPottimizzazione e i protocolli di trasferimento ottimizzati, ma una soluzione consiste nell'utilizzare User Datagram Protocol (). UDP

Anti-pattern comuni:

- Si utilizza TCP per tutti i carichi di lavoro indipendentemente dai requisiti di prestazioni.

Vantaggi dell'adozione di questa best practice: la verifica del protocollo adeguato per la comunicazione tra utenti e componenti del carico di lavoro contribuisce a migliorare l'esperienza utente complessiva per le applicazioni. Ad esempio, la modalità senza connessione UDP consente l'alta velocità, ma non offre ritrasmissione o alta affidabilità. TCPè un protocollo completo, ma richiede un sovraccarico maggiore per l'elaborazione dei pacchetti.

Livello di rischio associato se questa best practice non fosse adottata: medio

Guida all'implementazione

Se hai la possibilità di scegliere protocolli diversi per la tua applicazione e hai esperienza in questo campo, ottimizza l'applicazione e l'esperienza dell'utente finale utilizzando un protocollo diverso. Tieni conto che questo approccio presenta notevoli difficoltà e dovrebbe essere tentato solo dopo l'ottimizzazione dell'applicazione in altri modi.

Un aspetto principale per il miglioramento delle prestazioni del tuo carico di lavoro consiste nell'identificare i requisiti in termini di latenza e throughput, quindi scegliere i protocolli di rete che ottimizzano le prestazioni.

Quando prendere in considerazione l'utilizzo TCP

TCPfornisce una consegna affidabile dei dati e può essere utilizzato per la comunicazione tra i componenti del carico di lavoro in cui l'affidabilità e la consegna garantita dei dati sono importanti. Molte applicazioni basate sul Web si basano su protocolli TCP basati, come HTTP eHTTPS, per aprire i TCP socket per la comunicazione tra i componenti dell'applicazione. Il trasferimento di dati tramite posta elettronica e file sono applicazioni comuni che utilizzano anch'esseTCP, in quanto si tratta di un meccanismo di trasferimento semplice e affidabile tra i componenti dell'applicazione. L'utilizzo di TLS with TCP può comportare un sovraccarico di comunicazione, con conseguente aumento della latenza e riduzione della velocità effettiva, ma offre anche il vantaggio della sicurezza. Il sovraccarico è dovuto perlopiù al processo di handshake, il cui completamento può richiedere diversi round trip. Al termine del processo di handshake, il sovraccarico dovuto alla crittografia e alla decrittografia dei dati è relativamente ridotto.

Quando prendere in considerazione l'utilizzo UDP

UDPè un connection-less-oriented protocollo ed è quindi adatto per applicazioni che richiedono una trasmissione rapida ed efficiente, come log, monitoraggio e dati VoIP. Inoltre, UDP se disponi di componenti per il carico di lavoro, valuta la possibilità di utilizzare componenti che rispondono a piccole richieste provenienti da un gran numero di client per garantire prestazioni ottimali del carico di lavoro. Datagram Transport Layer Security (DTLS) è l'UDPequivalente di Transport Layer Security (). TLS Quando si utilizza DTLS withUDP, il sovraccarico deriva dalla crittografia e dalla decrittografia dei dati, poiché il processo di handshake è semplificato. DTLSaggiunge inoltre un piccolo sovraccarico ai UDP pacchetti, in quanto include campi aggiuntivi per indicare i parametri di sicurezza e rilevare eventuali manomissioni.

Quando prendere in considerazione l'utilizzo SRD

Scalable reliable datagram (SRD) è un protocollo di trasporto di rete ottimizzato per carichi di lavoro ad alto throughput grazie alla sua capacità di bilanciare il carico del traffico su più percorsi e ripristinare rapidamente il sistema in caso di interruzioni di pacchetti o errori di collegamento. SRDè quindi utilizzato al meglio per carichi di lavoro di elaborazione ad alte prestazioni (HPC) che richiedono un throughput elevato e una comunicazione a bassa latenza tra i nodi di elaborazione. Possono essere incluse attività di elaborazione in parallelo come la simulazione, la modellazione e l'analisi dei dati che implicano il trasferimento di grandi quantità di dati tra nodi.

Passaggi dell'implementazione

- Utilizzare i servizi AWS Global Accelerator e AWS Transfer Family per migliorare il throughput delle applicazioni di trasferimento file online. Il AWS Global Accelerator servizio consente di ridurre la latenza tra i dispositivi client e il carico di lavoro su. AWS Con AWS Transfer Family, puoi utilizzare protocolli TCP basati come Secure Shell File Transfer Protocol (SFTP) e File Transfer Protocol over SSL (FTPS) per scalare e gestire in modo sicuro i trasferimenti di file verso AWS i servizi di archiviazione.
- Utilizza la latenza di rete per determinare se TCP è appropriata per la comunicazione tra i componenti del carico di lavoro. Se la latenza di rete tra l'applicazione client e il server è elevata, l'handshake TCP a tre vie può richiedere del tempo, con un conseguente impatto sulla reattività dell'applicazione. Metriche come time to first byte (TTFB) e round-trip time (RTT) possono essere utilizzate per misurare la latenza di rete. Se il tuo carico di lavoro fornisce contenuti dinamici agli utenti, prendi in considerazione l'utilizzo di Amazon CloudFront, che stabilisce una connessione permanente a ciascuna origine per i contenuti dinamici per rimuovere i tempi di configurazione della connessione che altrimenti rallenterebbero ogni richiesta del client.
- L'utilizzo TLS con TCP o UDP può comportare un aumento della latenza e una riduzione del throughput per il carico di lavoro a causa dell'impatto della crittografia e della decrittografia. Per tali carichi di lavoro, prendi in considerazione l'opzioneSSL/TLSoffloading su Elastic Load Balancing per migliorare le prestazioni del carico di lavoro consentendo al sistema di bilanciamento del carico di SSL gestire il processo di crittografia e decrittografiaTLS/anziché affidarlo alle istanze di backend. Questo può aiutare a ridurre l'CPUutilizzo delle istanze di backend, migliorando le prestazioni e aumentando la capacità.
- Utilizza Network Load Balancer (NLB) per implementare servizi basati sul UDP protocollo, come autenticazione e autorizzazione, registrazione, DNS IoT e streaming multimediale, per migliorare le prestazioni e l'affidabilità del carico di lavoro. NLBDistribuisce il UDP traffico in entrata su più destinazioni, consentendoti di scalare il carico di lavoro orizzontalmente, aumentare la capacità e ridurre il sovraccarico di un singolo target.

- Per i carichi di lavoro High Performance Computing (HPC), prendi in considerazione l'utilizzo della funzionalità Elastic Network Adapter (ENA) Express che utilizza il SRD protocollo per migliorare le prestazioni di rete fornendo una maggiore larghezza di banda a flusso singolo (25 Gbps) e una latenza di coda inferiore (99,9 percentile) per il traffico di rete tra le istanze. EC2
- Utilizza Application Load Balancer (ALB) per indirizzare e bilanciare il carico del traffico g RPC (Remote Procedure Calls) tra i componenti del carico di lavoro o tra RPC i client e i servizi. g RPC utilizza il protocollo TCP basato su HTTP /2 per il trasporto e offre vantaggi in termini di prestazioni come un ingombro di rete più leggero, compressione, serializzazione binaria efficiente, supporto per numerose lingue e streaming bidirezionale.

Risorse

Documenti correlati:

- Come indirizzare il traffico verso Kubernetes UDP
- Application Load Balancer
- EC2Rete avanzata su Linux
- EC2Rete avanzata su Windows
- EC2Gruppi di collocamento
- Abilitazione di reti avanzate con Elastic Network Adapter (ENA) su istanze Linux
- Network Load Balancer
- Prodotti di rete con AWS
- Transitioning to Latency-Based Routing in Amazon Route 53
- VPCEndpoint

Video correlati:

- AWS re:Invent 2022 — Scalabilità delle prestazioni di rete sulle istanze Amazon Elastic Compute Cloud di nuova generazione
- AWS re:Invent 2022 — Fondamenti del networking delle applicazioni

Esempi correlati:

- AWS Transit Gateway e soluzioni di sicurezza scalabili
- Workshop sulle reti AWS

PERF04-BP06 Scegli la posizione del carico di lavoro in base ai requisiti di rete

Valuta le opzioni per il posizionamento delle risorse in modo da diminuire la latenza di rete e migliorare il throughput, fornendo un'esperienza utente ottimale attraverso la riduzione dei tempi di caricamento delle pagine e di trasferimento dei dati.

Anti-pattern comuni:

- Consolidamento di tutte le risorse del carico di lavoro in un'unica posizione geografica.
- Scelta della regione più vicina alla propria posizione, ma non al carico di lavoro dell'utente finale.

Vantaggi dell'adozione di questa best practice: l'esperienza utente è fortemente condizionata dalla latenza tra utente e applicazione. Utilizzando una rete globale appropriata Regioni AWS e AWS privata, è possibile ridurre la latenza e offrire un'esperienza migliore agli utenti remoti.

Livello di rischio associato se questa best practice non fosse adottata: medio

Guida all'implementazione

Le risorse, come le EC2 istanze Amazon, vengono collocate nelle Availability Zones within Regioni AWS, AWS Local Zones o AWS Wavelengthnelle zone. AWS Outposts La scelta della posizione influisce su latenza di rete e throughput dall'ubicazione di un utente specifico. I servizi edge come Amazon CloudFront AWS Global Acceleratorpossono essere utilizzati anche per migliorare le prestazioni di rete memorizzando nella cache i contenuti nelle edge location o fornendo agli utenti un percorso ottimale per il carico di lavoro attraverso la rete AWS globale.

Amazon EC2 fornisce gruppi di collocamento per il networking. Un gruppo di collocazione è un raggruppamento logico di istanze per ridurre la latenza. L'utilizzo di gruppi di collocamento con tipi di istanze supportati e un Elastic Network Adapter (ENA) consente ai carichi di lavoro di partecipare a una rete a 25 Gbps a bassa latenza e con jitter ridotto. I gruppi di collocazione sono consigliati per i carichi di lavoro che traggono beneficio da reti a bassa latenza, throughput di rete elevato o entrambi.

I servizi sensibili alla latenza vengono forniti nelle sedi periferiche utilizzando una rete AWS globale, come Amazon. CloudFront Queste edge location forniscono in genere servizi come Content Delivery Network (CDN) e Domain Name System (). DNS Disponendo di questi servizi all'edge, i carichi di lavoro possono rispondere con bassa latenza alle richieste di contenuto o DNS risoluzione. Inoltre, possono offrire servizi geografici come la geotargetizzazione dei contenuti (ossia fornire contenuti diversi in base alla posizione dell'utente finale) o l'instradamento basato sulla latenza, per indirizzare gli utenti alla regione più vicina (latenza minima).

Usa i servizi edge per ridurre la latenza e abilitare la memorizzazione nella cache dei contenuti. Configura correttamente il controllo della cache per entrambi DNS eHTTP/HTTPSper ottenere il massimo vantaggio da questi approcci.

Passaggi dell'implementazione

- Acquisisci informazioni sul traffico IP in entrata e in uscita dalle interfacce di rete.
  - Registrazione del traffico IP utilizzando VPC Flow Logs
  - Come viene preservato l'indirizzo IP del client in AWS Global Accelerator
- Analizza i modelli di accesso alla rete nel tuo carico di lavoro per capire come gli utenti usano la tua applicazione.
  - Utilizza strumenti di monitoraggio, come Amazon CloudWatch e AWS CloudTrail, per raccogliere dati sulle attività di rete.
  - Analizza i dati per identificare il modello di accesso alla rete.
- Seleziona regioni appropriate per l'implementazione del carico di lavoro in base ai seguenti elementi chiave:
  - Ubicazione dei dati per le applicazioni a uso intensivo di dati, ad esempio applicazioni di big data e machine learning, il codice dell'applicazione dovrebbe essere eseguito il più vicino possibile ai dati.
  - Ubicazione degli utenti: per le applicazioni rivolte agli utenti, scegli una regione o più regioni vicine agli utenti del carico di lavoro.
  - Altri vincoli: prendi in considerazione vincoli come costi e conformità, come illustrato in What to Consider when Selecting a Region for your Workloads.
- Usa le zone locali AWS per eseguire carichi di lavoro come il rendering video. Le zone locali consentono di sfruttare i vantaggi derivanti dalla disponibilità di risorse di calcolo e archiviazione più vicine agli utenti finali.
- Usa AWS Outposts per carichi di lavoro che devono rimanere on-premises, ma vuoi che vengano eseguiti in modo ottimale con il resto degli altri carichi di lavoro in AWS.
- Applicazioni come lo streaming video in diretta ad alta risoluzione, l'audio ad alta fedeltà e la realtà aumentata o la realtà virtuale (AR/VR) richiedono dispositivi 5G. ultra-low-latency Per tali applicazioni, considera. AWS Wavelength AWS Wavelength incorpora servizi di AWS elaborazione e archiviazione nelle reti 5G, fornendo un'infrastruttura di edge computing mobile per lo sviluppo, l'implementazione e la scalabilità delle applicazioni. ultra-low-latency
- Usa la cache locale o le soluzioni di caching AWS per i dati di frequente utilizzo per migliorare le performance, ridurre lo spostamento dei dati e minimizzare l'impatto ambientale.

Servizio	Quando usare
Amazon CloudFront	Utilizzalo per memorizzare nella cache contenuti statici come immagini, script e video, nonché contenuti dinamici come API risposte o applicazioni web.
Amazon ElastiCache	Usalo per memorizzare nella cache i contenuti per le applicazioni Web.
DynamoDB Accelerator	Usalo per aggiungere accelerazione in memoria alle tabelle DynamoDB.

- Utilizza servizi in grado di supportarti nell'esecuzione del codice in posizioni più vicine agli utenti del carico di lavoro, come i seguenti:

Servizio	Quando usare
Lambda@Edge	Usalo per operazioni a uso intensivo di risorse di calcolo eseguite quando gli oggetti non si trovano nella cache.
CloudFront Funzioni Amazon	Utilizzalo per casi d'uso semplici come richieste HTTP (s) o manipolazioni di risposte che possono essere avviate da funzioni di breve durata.
AWS IoT Greengrass	Usale per eseguire la memorizzazione nella cache di risorse di calcolo, messaggistica e dati per i dispositivi connessi.

- Alcune applicazioni richiedono punti di ingresso fissi o prestazioni più elevate attraverso la riduzione della latenza di ricezione del primo byte e l'instabilità e l'aumento del throughput. Queste applicazioni possono trarre vantaggio da servizi di rete che forniscono indirizzi IP anycast statici e TCP terminazioni in postazioni periferiche. AWS Global Acceleratorpossono migliorare le prestazioni delle applicazioni fino al 60% e fornire un failover rapido per architetture multiregionali. AWS Global Accelerator fornisce indirizzi IP anycast statici che fungono da punto di ingresso fisso

per le applicazioni ospitate in una o più applicazioni. Regioni AWS Questi indirizzi IP consentono al traffico di entrare nella rete AWS globale il più vicino possibile agli utenti. AWS Global Accelerator riduce il tempo di configurazione iniziale della connessione stabilendo una TCP connessione tra il client e la AWS edge location più vicina al client. Rivedi l'utilizzo di AWS Global Accelerator per migliorare le prestazioni dei tuoi UDP carichi di lavoroTCP/e fornire un failover rapido per architetture multiregionali.

Risorse

Best practice correlate:

- COST07-BP02 Implementazione delle regioni in base ai costi
- COST08-BP03 Implementazione di servizi per ridurre i costi di trasferimento dei dati
- REL10-BP01 Implementa il carico di lavoro in più sedi
- REL10-BP02 Seleziona le posizioni appropriate per l'implementazione in più sedi
- SUS01-BP01 Scegli la regione in base ai requisiti aziendali e agli obiettivi di sostenibilità
- SUS02-BP04 Ottimizza il posizionamento geografico dei carichi di lavoro in base ai requisiti di rete
- SUS04-BP07 Riduci al minimo lo spostamento dei dati tra le reti

Documenti correlati:

- AWS Infrastruttura globale
- AWS Local Zones e AWS Outposts scelta della tecnologia giusta per il tuo carico di lavoro edge
- Placement groups
- AWS Local Zones
- AWS Outposts
- AWS Wavelength
- Amazon CloudFront
- AWS Global Accelerator
- AWS Direct Connect
- AWS Site-to-Site VPN
- Amazon Route 53

Video correlati:

- AWS Video esplicativo su Local Zones
- AWS Outposts: Overview and How it Works
- AWS re:Invent 2023 - Una strategia di migrazione per carichi di lavoro edge e locali
- AWS re:Invent 2021 -: Portare l'esperienza in sede AWS OutpostsAWS
- AWS re:Invent 2020:: Esegui app con latenza AWS Wavelength ultra bassa su 5G Edge
- AWS re:Invent 2022 - AWS Local Zones: creazione di applicazioni per un edge distribuito
- AWS re:Invent 2021 - Creazione di siti Web a bassa latenza con Amazon CloudFront
- AWS re:Invent 2022 - Migliora le prestazioni e la disponibilità con AWS Global Accelerator
- AWS re:Invent 2022 - Costruisci la tua rete WAN utilizzando AWS
- AWS re:Invent 2020: gestione globale del traffico con Amazon Route 53

Esempi correlati:

- AWS Global Accelerator Workshop sul routing personalizzato
- Handling Rewrites and Redirects using Edge Functions

PERF04-BP07 Ottimizza la configurazione di rete in base a metriche

Usa i dati raccolti e analizzati per prendere decisioni informate riguardo l'ottimizzazione della configurazione della tua rete.

Anti-pattern comuni:

- Si ritiene che tutti i problemi relativi alle prestazioni siano correlati all'applicazione.
- Verifica delle prestazioni di rete solo da una posizione vicina a quella in cui è stato distribuito il carico di lavoro.
- Uso di configurazioni predefinite per tutti i servizi di rete.
- Provisioning in eccesso di risorse di rete per fornire capacità sufficiente.

Vantaggi dell'adozione di questa best practice: la raccolta delle metriche necessarie per la rete AWS e l'implementazione di strumenti di monitoraggio di rete permettono di identificare le prestazioni di rete e ottimizzare le configurazioni di rete.

Livello di rischio associato se questa best practice non fosse adottata: basso

Guida all'implementazione

Il monitoraggio del traffico da e versoVPCs, le sottoreti o le interfacce di rete è fondamentale per capire come utilizzare le risorse di rete e ottimizzare le configurazioni di rete. AWS Utilizzando i seguenti strumenti AWS di rete, è possibile esaminare ulteriormente le informazioni sull'utilizzo del traffico, sull'accesso alla rete e sui registri.

Passaggi dell'implementazione

- Identifica le metriche prestazionali chiave, come la latenza o la perdita di pacchetti, da raccogliere. AWS fornisce diversi strumenti che possono aiutarti a raccogliere queste metriche. Usando i seguenti strumenti, puoi esaminare ulteriormente le informazioni sull'utilizzo del traffico, sull'accesso alla rete e sui log:

AWS strumento	Dove usarlo
Gestione indirizzi VPC IP di Amazon.	IPAMUtilizzalo per pianificare, tracciare e monitorare gli indirizzi IP per i tuoi carichi di lavoro AWS e quelli locali. Si tratta di una best practice per ottimizzare l'utilizzo e l'allocazione degli indirizzi IP.
VPCRegistri di flusso	Usa VPC Flow Logs per acquisire informazi oni dettagliate sul traffico da e verso le interfacce di rete del tuo. VPCs Con VPC Flow Logs, puoi diagnosticare regole di gruppo di sicurezza eccessivamente restrittive o permissive e determinare la direzione del traffico da e verso le interfacce di rete.
AWS Transit Gateway Log di flusso	Utilizzate AWS Transit Gateway Flow Logs per acquisire informazioni sul traffico IP in entrata e in uscita dai gateway di transito.
DNSregistrazione delle interrogazioni	Registra le informazioni sulle DNS interroga zioni pubbliche o private ricevute da Route 53. Con DNS i log, è possibile ottimizzare le DNS

AWS strumento	Dove usarlo
	configurazioni comprendendo il dominio o il sottodominio richiesto o le EDGE posizioni di Route 53 che hanno risposto alle query. DNS
Reachability Analyzer	Con Reachability Analyzer puoi analizzare la raggiungibilità della rete ed eseguirne il debug. Reachability Analyzer è uno strumento di analisi della configurazione che consente di eseguire test di connettività tra una risorsa di origine e una risorsa di destinazione nel proprio. VPCs Lo strumento in questione consente di verificare la corrispondenza fra configurazione e connettività desiderata.
Strumento di analisi degli accessi alla rete	È possibile utilizzare lo Strumento di analisi degli accessi alla rete per comprendere l'accesso di rete alle risorse. Puoi usare lo Strumento di analisi degli accessi alla rete per specificare i requisiti di accesso alla rete e identificare i potenziali percorsi di rete che non li soddisfano. Ottimizzando la configura zione di rete corrispondente, puoi determinare e verificare lo stato della rete e indicare se la rete su AWS soddisfa i requisiti di conformità.

AWS strumento	Dove usarlo
Amazon CloudWatch	Usa Amazon CloudWatch e attiva i parametri appropriati per le opzioni di rete. Assicurati di scegliere le metriche di rete corrette per il carico di lavoro. Ad esempio, puoi attivare le metriche per VPC Network Address Usage, VPC NAT Gateway AWS Transit Gateway, VPN tunnel AWS Network Firewall, Elastic Load Balancing e. AWS Direct Connect Il monitoraggio continuo delle metriche è una procedura utile per osservare e identificare lo stato e l'utilizzo della rete che semplifica l'ottimizzazione della configurazione di rete in base alle osservazioni.
AWS Network Manager	Utilizzando AWS Network Manager, è possibile monitorare le prestazioni in tempo reale e cronologiche della rete AWS globale per scopi operativi e di pianificazione. Network Manager fornisce la latenza di rete aggregata tra le zone di disponibilità Regioni AWS e all'interno di ciascuna zona di disponibilità, consentendovi di comprendere meglio in che modo le prestazioni delle applicazioni si relazionano con le prestazioni della rete sottostante. AWS
Amazon CloudWatch RUM	Usa Amazon CloudWatch RUM per raccoglie re le metriche che ti forniscono le informazioni che ti aiutano a identificare, comprendere e migliorare l'esperienza utente.

- Identifica i principali oratori e i modelli di traffico delle applicazioni utilizzando VPC e AWS Transit Gateway Flow Logs.

- Valuta e ottimizza la tua attuale architettura di reteVPCs, inclusi sottoreti e routing. Ad esempio, potete valutare la diversità del VPC peering o AWS Transit Gateway aiutarvi a migliorare il networking nella vostra architettura.
- Valuta i percorsi di instradamento nella tua rete per verificare che venga sempre utilizzato il percorso più breve tra le destinazioni. Lo Strumento di analisi degli accessi alla rete è utile in questa operazione.

Risorse

Documenti correlati:

- DNSRegistrazione pubblica delle interrogazioni
- Che cos'èIPAM?
- What is Reachability Analyzer?
- What is Network Access Analyzer?
- CloudWatchmetriche per te VPCs
- Ottimizza le prestazioni e riduci i costi per l'analisi di rete con VPC Flow Logs in formato Apache Parquet
- Monitoraggio delle tue reti globali e principali con i CloudWatch parametri di Amazon
- Continuously monitor network traffic and resources

Video correlati:

- AWS re:Invent 2023: guida per sviluppatori al cloud networking
- AWS re:Invent 2023 — Pronti per il futuro? Designing networks for growth and flexibility
- AWS re:Invent 2023 — Design avanzati e nuove funzionalità VPC
- AWS re:Invent 2022 — Approfondisci l'infrastruttura di rete AWS
- AWS re:Invent 2020 — Le migliori pratiche e suggerimenti per il networking con il Well-Architected Framework AWS
- AWS re:Invent 2020 — Monitoraggio e risoluzione dei problemi del traffico di rete

Esempi correlati:

- Workshop sulle reti AWS

- AWS Network Monitoring
- Osservazione e diagnosi della rete su AWS
- Individuazione e risoluzione degli errori di configurazione della rete su AWS

Processo e cultura

Questions

- PERF5. In che modo le pratiche e la cultura dell'organizzazione contribuiscono all'efficienza delle prestazioni nel carico di lavoro?

PERF5. In che modo le pratiche e la cultura dell'organizzazione contribuiscono all'efficienza delle prestazioni nel carico di lavoro?

Durante la fase di progettazione dei carichi di lavoro, esistono principi e pratiche che è possibile adottare per gestire al meglio carichi di lavoro cloud efficienti e ad alte prestazioni. Per adottare una cultura che promuova l'efficienza delle prestazioni dei carichi di lavoro cloud, prendi in considerazione questi principi e pratiche fondamentali:

Best practice

- PERF05-BP01 Stabilire indicatori chiave di prestazione (KPIs) per misurare lo stato e le prestazioni del carico di lavoro
- PERF05-BP02 Utilizza soluzioni di monitoraggio per comprendere le aree in cui le prestazioni sono più critiche
- PERF05-BP03 Definire un processo per migliorare le prestazioni del carico di lavoro
- PERF05-BP04 Load Esegui un test del tuo carico di lavoro
- PERF05-BP05 Usa l'automazione per risolvere in modo proattivo i problemi relativi alle prestazioni
- PERF05-BP06 Conserva il carico di lavoro e i servizi up-to-date
- PERF05-BP07 Rivedi le metriche a intervalli regolari

PERF05-BP01 Stabilire indicatori chiave di prestazione (KPIs) per misurare lo stato e le prestazioni del carico di lavoro

Identifica quelli KPIs che misurano quantitativamente e qualitativamente le prestazioni del carico di lavoro. KPIsti aiutano a misurare lo stato e le prestazioni di un carico di lavoro correlato a un obiettivo aziendale.

Anti-pattern comuni:

- Monitoraggio dei parametri a livello di sistema solo per avere una visione del carico di lavoro e mancata valutazione degli impatti aziendali di tali parametri.
- Partite dal presupposto che i vostri dati KPIs siano già stati pubblicati e condivisi come dati metrici standard.
- Non definisci un valore quantitativo e misurabile. KPI
- Non siete in linea KPIs con gli obiettivi o le strategie aziendali.

Vantaggi derivanti dall'adozione di questa best practice: l'identificazione di specifiche KPIs che rappresentino lo stato e le prestazioni del carico di lavoro aiuta ad allineare i team sulle loro priorità e a definire risultati aziendali di successo. La condivisione di tali metriche con tutti i reparti fornisce visibilità e allineamento su soglie, aspettative e impatto aziendale.

Livello di rischio associato se questa best practice non fosse adottata: elevato

Guida all'implementazione

KPIsconsentire ai team aziendali e tecnici di allinearsi sulla misurazione degli obiettivi e delle strategie e sul modo in cui questi fattori si combinano per produrre risultati aziendali. Ad esempio, il carico di lavoro di un sito Web può utilizzare il tempo di caricamento della pagina come indicazione delle prestazioni complessive. Questa metrica sarebbe uno dei vari punti dati che misurano l'esperienza dell'utente. Oltre a identificare le soglie di tempo di caricamento della pagina, occorre documentare il risultato atteso o il rischio aziendale in caso di mancato raggiungimento delle prestazioni ideali. Un lungo tempo di caricamento della pagina si ripercuote direttamente sugli utenti finali, peggiora la loro esperienza d'uso e può portare a una perdita di clienti. Quando definisci le tue KPI soglie, combina i benchmark di settore e le aspettative degli utenti finali. Ad esempio, se l'attuale benchmark di settore è una pagina Web che viene caricata entro un periodo di tempo di due secondi, ma gli utenti finali si aspettano che una pagina Web venga caricata entro un periodo di tempo di un secondo, è necessario prendere in considerazione entrambi questi dati quando si stabilisce il. KPI

Il team deve valutare il carico di lavoro KPIs utilizzando dati granulari e storici in tempo reale come riferimento e creare dashboard che eseguano calcoli metrici sui dati per ricavare informazioni operative e di utilizzo. KPI KPIsdevono essere documentati e includere soglie che supportano gli obiettivi e le strategie aziendali e devono essere mappati alle metriche monitorate. KPIsdevono essere riesaminate quando gli obiettivi aziendali, le strategie o i requisiti degli utenti finali cambiano.

Passaggi dell'implementazione

- Identifica le parti interessate: identifica e documenta le principali parti interessate aziendali, compresi i team di sviluppo e operativi.
- Definisci gli obiettivi: collabora con queste parti interessate per definire e documentare gli obiettivi del carico di lavoro. Considera gli aspetti critici relativi alle prestazioni dei carichi di lavoro, come il throughput, i tempi di risposta e i costi, nonché gli obiettivi aziendali, come la soddisfazione degli utenti.
- Esamina le best practice del settore: esamina le best practice del settore per identificare le soluzioni pertinenti in KPIs linea con gli obiettivi del carico di lavoro.
- Individua le metriche: identifica le metriche in linea con gli obiettivi del carico di lavoro e in grado di aiutarti a misurare prestazioni e obiettivi aziendali. Stabilisci KPIs in base a queste metriche. ad esempio le misurazioni del tempo medio di risposta o del numero di utenti simultanei.
- Definisci e documentaKPIs: utilizza le migliori pratiche del settore e gli obiettivi del carico di lavoro per fissare obiettivi per il tuo carico di lavoro. KPI Utilizza queste informazioni per impostare KPI soglie di gravità o livello di allarme. Identifica e documenta il rischio e l'impatto di un problema KPI non soddisfatto.
- Implementa il monitoraggio: utilizza strumenti di monitoraggio come Amazon CloudWatch o AWS Configper raccogliere metriche e misurareKPIs.
- Comunicazione visivaKPIs: utilizza strumenti di dashboard come Amazon QuickSight per visualizzare e comunicare KPIs con le parti interessate.
- Analizza e ottimizza: rivedi e analizza regolarmente KPIs per identificare le aree del tuo carico di lavoro che devono essere migliorate. Collabora con le parti interessate per implementare tali miglioramenti.
- Rivedi e perfeziona: rivedi regolarmente le metriche e KPIs valutane l'efficacia, soprattutto quando gli obiettivi aziendali o le prestazioni del carico di lavoro cambiano.

Risorse

Documenti correlati:

- CloudWatchdocumentazione
- Monitoraggio, registrazione e prestazioni s AWS Partner
- AWS strumenti di osservabilità
- L'importanza degli indicatori chiave di prestazione (KPIs) per le migrazioni al cloud su larga scala
- Come monitorare l'ottimizzazione dei costi KPIs con la Dashboard KPI
- Documentazione di X-Ray
- Utilizzo delle CloudWatch dashboard di Amazon
- Amazon QuickSight KPIs

Video correlati:

- AWS re:Invent 2023 - Ottimizza costi e prestazioni e monitora i progressi verso la mitigazione
- AWS re:Invent 2023 - Gestisci gli eventi del ciclo di vita delle risorse su larga scala con AWS Health
- AWS re:Invent 2023 - Prestazioni ed efficienza su Pinterest: ottimizzazione delle istanze più recenti
- AWS re:Invent 2022 - ottimizzazione: misure attuabili per risultati immediati AWS
- AWS re:Invent 2023 - Costruire un'efficace strategia di osservabilità
- AWS Summit SF 2022 - Osservabilità completa e monitoraggio delle applicazioni con AWS
- AWS re:Invent 2023 - Scalabilità per i primi 10 milioni di utenti AWS
- AWS re:Invent 2022 - In che modo Amazon utilizza metriche migliori per migliorare le prestazioni del sito Web
- Creazione di una strategia di metrica efficace per la tua azienda | Eventi AWS

Esempi correlati:

- Creazione di una dashboard con Amazon QuickSight

PERF05-BP02 Utilizza soluzioni di monitoraggio per comprendere le aree in cui le prestazioni sono più critiche

Comprendi e identifica le aree in cui l'aumento delle prestazioni del carico di lavoro determinerà un impatto positivo sull'efficienza o sull'esperienza del cliente. Ad esempio, un sito web che ha una grande quantità di interazione con i clienti può trarre vantaggio dall'utilizzo dei servizi edge per spostare la distribuzione di contenuti più vicino ai clienti.

Anti-pattern comuni:

- Si presume che le metriche di elaborazione standard come l'CPUutilizzo o la pressione della memoria siano sufficienti per individuare problemi di prestazioni.
- Utilizzo solo dei parametri predefiniti registrati dal software di monitoraggio selezionato.
- Revisione dei parametri solo quando c'è un problema.

Vantaggi derivanti dall'adozione di questa best practice: la comprensione delle aree critiche delle prestazioni aiuta i proprietari dei carichi di lavoro a monitorare KPIs e dare priorità ai miglioramenti ad alto impatto.

Livello di rischio associato se questa best practice non fosse adottata: elevato

Guida all'implementazione

Imposta il end-to-end tracciamento per identificare i modelli di traffico, la latenza e le aree critiche di prestazioni. Monitora gli schemi di accesso ai dati per query lente o dati scarsamente frammentati e partizionati. Identifica le aree vincolate del carico di lavoro utilizzando test o monitoraggio del carico.

Aumenta l'efficienza delle prestazioni esaminando l'architettura, gli schemi di traffico e gli schemi di accesso ai dati e identifica la latenza e i tempi di elaborazione. Identifica i potenziali colli di bottiglia che potrebbero influire sull'esperienza del cliente man mano che il carico di lavoro aumenta. Dopo aver identificato queste aree, individua quale soluzione puoi implementare per evitare tali problemi di prestazioni.

Passaggi dell'implementazione

- Imposta il end-to-end monitoraggio per acquisire tutti i componenti e le metriche del carico di lavoro. Ecco alcuni esempi di soluzioni di monitoraggio su. AWS

Servizio	Dove usarlo
Amazon CloudWatch Real-User Monitoring () RUM	Per acquisire i parametri delle prestazioni delle applicazioni da sessioni lato client e frontend di utenti reali.
AWS X-Ray	Per tenere traccia del traffico nei livelli dell'applicazione e identificare la latenza tra componenti e dipendenze. Utilizza le mappe

Servizio	Dove usarlo
	del servizio X-Ray per osservare le relazioni e la latenza tra i componenti del carico di lavoro.
Informazioni dettagliate sulle prestazioni del servizio Amazon Relational Database	Per osservare i parametri delle prestazioni del database e identificare le prestazioni da migliorare.
Monitoraggio RDS avanzato di Amazon	Per osservare i parametri delle prestazioni del sistema operativo del database.
Amazon DevOps Guru	Per rilevare modelli operativi anomali in modo da poter identificare i problemi operativi prima che abbiano un impatto sui clienti.

- Esegui i test per generare parametri, identificare schemi di traffico, colli di bottiglia e aree con prestazioni critiche. Ecco alcuni esempi di come eseguire i test:
  - Configura CloudWatchSynthetic Canaries per imitare le attività degli utenti basate su browser in modo programmatico utilizzando cron job Linux o espressioni di frequenza per generare metriche coerenti nel tempo.
  - Usa la soluzione Test di carico distribuito di AWS per generare picchi di traffico o testare il carico di lavoro al tasso di crescita previsto.
- Valuta parametri e dati di telemetria per identificare le aree critiche delle prestazioni. Esamina queste aree con il tuo team per determinare il monitoraggio e le soluzioni per evitare i colli di bottiglia.
- Sperimenta i miglioramenti delle prestazioni e valuta tali modifiche con i dati. Ad esempio, puoi usare CloudWatchEvidently per testare nuovi miglioramenti e impatti prestazionali sul tuo carico di lavoro.

Risorse

Documenti correlati:

- Cosa c'è di nuovo in AWS Observability a re:Invent 2023
- Amazon Builders' Library
- Documentazione di X-Ray

- Amazon CloudWatch RUM
- Amazon DevOps Guru

Video correlati:

- AWS re:Invent 2023 - [LAUNCH] Monitoraggio delle applicazioni per carichi di lavoro moderni
- AWS re:Invent 2023 - Implementazione dell'osservabilità delle applicazioni
- AWS re:Invent 2023 - Costruire una strategia di osservabilità efficace
- AWS Summit SF 2022 - Osservabilità completa e monitoraggio delle applicazioni con AWS
- AWS Re:Invent 2022 - AWS ottimizzazione: passaggi attuabili per risultati immediati
- AWS re:Invent 2022 - The Amazon Builders' Library: 25 anni di eccellenza operativa di Amazon
- AWS re:Invent 2022 - In che modo Amazon utilizza metriche migliori per migliorare le prestazioni del sito Web
- Monitoraggio visivo delle applicazioni con Amazon CloudWatch Synthetics

Esempi correlati:

- Misura il tempo di caricamento della pagina con Amazon CloudWatch Synthetics
- Client CloudWatch RUM Web Amazon
- X-Ray SDK per Python
- Test di carico distribuito su AWS

PERF05-BP03 Definire un processo per migliorare le prestazioni del carico di lavoro

Definisci un processo per valutare i nuovi servizi, i modelli di progettazione, i tipi di risorse e le configurazioni man mano che diventano disponibili. Ad esempio, esegui test delle prestazioni esistenti sulle nuove offerte di istanze per determinare il loro potenziale per migliorare il carico di lavoro.

Anti-pattern comuni:

- Si ritiene che l'architettura corrente diventi statica e non venga aggiornata nel corso del tempo.
- Introduzione di modifiche all'architettura nel tempo senza dei parametri che le giustifichino.

Vantaggi dell'adozione di questa best practice: definire un processo per apportare modifiche all'architettura consente ai dati raccolti di influenzare la progettazione del carico di lavoro nel corso del tempo.

Livello di rischio associato se questa best practice non fosse adottata: medio

Guida all'implementazione

Le prestazioni del carico di lavoro presentano alcuni vincoli principali. Documentali, in modo da sapere quali tipi di innovazione potrebbero migliorare le prestazioni del carico di lavoro. Utilizza queste informazioni quando vieni a conoscenza di nuovi servizi o tecnologie, man mano che si rendono disponibili, in modo da identificare le soluzioni per ovviare ai vincoli o ai colli di bottiglia.

Determina i principali vincoli riguardanti le prestazioni del carico di lavoro. Documenta i vincoli prestazionali del carico di lavoro in modo da sapere quali tipi di innovazione potrebbero migliorare le prestazioni del carico di lavoro.

Passaggi dell'implementazione

- IdentificazioneKPIs: identifica le prestazioni del carico di lavoro KPIs come indicato nella tabella di base del carico di lavoro. PERF05-BP01 Stabilire indicatori chiave di prestazione (KPIs) per misurare lo stato e le prestazioni del carico di lavoro
- Implementa il monitoraggio: utilizza strumenti di AWS osservabilità per raccogliere metriche e misurare le prestazioni. KPIs
- Effettua analisi: conduci anali approfondite per individuare le aree (come la configurazione e il codice applicativo) del carico di lavoro con prestazioni insufficienti, come indicato in PERF05-BP02 Utilizza soluzioni di monitoraggio per comprendere le aree in cui le prestazioni sono più critiche. Usa i tuoi strumenti di analisi e prestazioni per individuare la strategia di miglioramento delle prestazioni.
- Convalida i miglioramenti: utilizza gli ambienti sandbox o di preproduzione per convalidare l'efficacia della strategia di miglioramento.
- Implementa le modifiche: implementa le modifiche nella produzione e monitora in modo continuo le prestazioni del carico di lavoro. Documenta i miglioramenti e comunica i risultati alle parti interessate.
- Riesamina e perfeziona: rivedi con regolarità il processo di miglioramento delle prestazioni per individuare le aree di miglioramento.

Risorse

Documenti correlati:

- Blog AWS
- Cosa c'è di nuovo con AWS
- AWS Skill Builder

Video correlati:

- AWS re:Invent 2022 - Fornire architetture sostenibili e ad alte prestazioni
- AWS re:Invent 2023 - Ottimizza costi e prestazioni e monitora i progressi verso la mitigazione
- AWS re:Invent 2022 - AWS ottimizzazione: misure attuabili per risultati immediati
- AWS re:Invent 2022 - Ottimizza i tuoi carichi di lavoro seguendo le migliori pratiche AWS

Esempi correlati:

- AWS Github

PERF05-BP04 Load Esegui un test del tuo carico di lavoro

Esegui il test del carico di lavoro per verificare che sia in grado di gestire il carico di produzione e individuare eventuali colli di bottiglia nelle prestazioni.

Anti-pattern comuni:

- Test delle singole parti del carico di lavoro, ma non dell'intero carico di lavoro.
- Test di carico eseguito su un'infrastruttura diversa dall'ambiente di produzione.
- Test di carico eseguiti solo per il carico previsto e non oltre, per prevedere dove si potrebbero riscontrare problemi futuri.
- Esegui test di carico senza consultare la Amazon EC2 Testing Policy e inviare un modulo di invio di eventi simulati. Ciò comporta la mancata esecuzione del test, in quanto sembra un evento. denial-of-service

Vantaggi dell'adozione di questa best practice: misurando le prestazioni in un test di carico, potrai vedere dove avrà luogo l'impatto con l'aumento del carico. In questo modo puoi anticipare le modifiche necessarie prima che influiscano sul carico di lavoro.

Livello di rischio associato se questa best practice non fosse adottata: basso

Guida all'implementazione

Il test di carico nel cloud è un processo volto a misurare le prestazioni del carico di lavoro in condizioni realistiche e con il carico degli utenti previsto. Questo processo prevede il provisioning di un ambiente cloud simile a quello di produzione, l'utilizzo di strumenti di test di carico per generare il carico e l'analisi dei parametri per valutare la capacità del carico di lavoro di gestire un carico realistico. Occorre eseguire i test di carico tramite versioni sintetiche o purificate dei dati di produzione (rimuovendo le informazioni sensibili o che permettono l'identificazione degli utenti). Eseguite automaticamente i test di carico come parte della vostra pipeline di distribuzione e confrontate i risultati con soglie e soglie predefiniteKPIs. Questo processo ti consente di ottenere le prestazioni richieste.

Passaggi dell'implementazione

- Definisci gli obiettivi dei test: individua gli aspetti in termini di prestazione del carico di lavoro da valutare, come il throughput e il tempo di risposta.
- Seleziona uno strumento di test: scegli e configura lo strumento di test più adatto al carico di lavoro.
- Configura l'ambiente: configura l'ambiente di test in base al tuo ambiente di produzione. Puoi utilizzare AWS i servizi per eseguire ambienti su scala di produzione per testare la tua architettura.
- Implementa il monitoraggio: utilizza strumenti di monitoraggio come Amazon CloudWatch per raccogliere metriche tra le risorse della tua architettura. Puoi anche raccogliere e pubblicare metriche personalizzate.
- Definisci gli scenari definisci scenari e parametri del test di carico (come la durata del test e il numero di utenti).
- Esegui test di carico: effettua scenari di test su vasta scala. Approfittane Cloud AWS per testare il tuo carico di lavoro e scoprire dove non riesce a scalare o se è scalabile in modo non lineare. Ad esempio, usa le istanze spot per generare carichi a costi ridotti e rilevare i colli di bottiglia prima che si verifichino in produzione.
- Analizza i risultati dei test: analizza i risultati per individuare colli di bottiglia delle prestazioni e aree di miglioramento.

- Documenta e condividi gli esiti: documenta esiti e raccomandazioni e crea report al riguardo. Condividi queste informazioni con le parti interessate per aiutarle a prendere decisioni informate sulle strategie di ottimizzazione delle prestazioni.
- Effettua iterazioni continue: esegui con regolarità i test di carico, specie dopo una modifica o un aggiornamento del sistema.

Risorse

Documenti correlati:

- Amazon CloudWatch RUM
- Amazon CloudWatch Synthetics
- Test di carico distribuito su AWS

Video correlati:

- AWS Summit ANZ 2023: accelera con fiducia grazie ai test di carico AWS distribuiti
- AWS re:Invent 2022: scalabile AWS per i primi 10 milioni di utenti
- Soluzione con AWS soluzioni: test di carico distribuiti

- AWS re:Invent 2021 - Ottimizza le applicazioni attraverso approfondimenti sugli utenti finali con Amazon CloudWatch RUM
- Demo di Amazon CloudWatch Synthetics

Esempi correlati:

- Test di carico distribuito su AWS

PERF05-BP05 Usa l'automazione per risolvere in modo proattivo i problemi relativi alle prestazioni

Utilizzate gli indicatori chiave di prestazione (KPIs), combinati con i sistemi di monitoraggio e avviso, per affrontare in modo proattivo i problemi relativi alle prestazioni.

Anti-pattern comuni:

- Solo il personale operativo è autorizzato ad apportare modifiche operative al carico di lavoro.

- Tutti gli allarmi giungono direttamente al team operativo senza alcuna correzione proattiva.

Vantaggi dell'adozione di questa best practice: la correzione proattiva delle azioni di allarme consente al personale di supporto di concentrarsi sugli elementi non attivabili in automatico. In questo modo, il personale operativo non viene sovraccaricato da tutti gli allarmi e si concentra, invece, solo sugli allarmi critici.

Livello di rischio associato se questa best practice non fosse adottata: basso

Guida all'implementazione

Laddove possibile, utilizza gli allarmi per attivare operazioni automatizzate per risolvere i problemi. Se non è possibile rispondere in modo automatizzato, inoltra l'allarme a chi può intervenire. Ad esempio, potreste disporre di un sistema in grado di prevedere i valori previsti degli indicatori chiave di prestazione (KPI) e di avvisare quando superano determinate soglie, oppure uno strumento in grado di arrestare o ripristinare automaticamente le implementazioni se non raggiungono i valori previsti. KPIs

Implementa processi che forniscono visibilità sulle prestazioni durante l'esecuzione del carico di lavoro. Crea pannelli di controllo del monitoraggio e stabilisci norme di riferimento per le aspettative in termini di prestazioni, per determinare se il carico di lavoro presenta prestazioni ottimali.

Passaggi dell'implementazione

- Identifica il flusso di correzione: individua e comprendi il problema delle prestazioni risolvibile automaticamente. Utilizza soluzioni di AWS monitoraggio come Amazon CloudWatch o AWS X-Ray per aiutarti a comprendere meglio la causa principale del problema.
- Definisci il processo di automazione: crea un processo di step-by-step riparazione che può essere utilizzato per risolvere automaticamente il problema.
- Configura l'evento di avvio: configura l'evento per l'avvio automatico del processo di risoluzione. Ad esempio, è possibile definire un trigger per riavviare automaticamente un'istanza quando raggiunge una determinata soglia di CPU utilizzo.
- Automatizza la riparazione: utilizza AWS servizi e tecnologie per automatizzare il processo di riparazione. Ad esempio, AWS Systems Manager Automation fornisce un modo sicuro e scalabile per automatizzare il processo di risoluzione. Assicurati di utilizzare la logica di risoluzione automatica per annullare le modifiche se non risolvono correttamente il problema.
- Testa il flusso di lavoro: esegui il test del processo di risoluzione automatizzato in un ambiente di preproduzione.

- Implementa il flusso di lavoro: implementa la risoluzione automatizzata nell'ambiente di produzione.
- Sviluppa un playbook: predisponi e documenta un playbook che delinei le fasi del piano di risoluzione, inclusi eventi di avvio, logica di risoluzione e azioni intraprese. Assicurati di fornire la giusta preparazione alle parti interessate in modo che possano rispondere efficacemente agli eventi di risoluzione automatizzati.
- Esamina e perfeziona: valuta con regolarità l'efficacia del flusso di lavoro di risoluzione automatizzato. Modifica gli eventi di avvio e la logica di risoluzione, se necessario.

Risorse

Documenti correlati:

- CloudWatchDocumentazione
- Partner per il monitoraggio, la registrazione e le prestazioni AWS Partner Network
- Documentazione di X-Ray
- Utilizzo di allarmi e azioni di allarme in CloudWatch
- Sviluppa una pratica di automazione del cloud per l'eccellenza operativa: le migliori pratiche di AWS Managed Services
- Automate your Amazon Redshift performance tuning with automatic table optimization

Video correlati:

- AWS re:Invent 2023 - Strategie per la scalabilità automatizzata, la correzione e l'autoguarigione intelligente
- AWS re:Invent 2023 - [] Monitoraggio delle applicazioni per carichi di lavoro moderni LAUNCH
- AWS re:Invent 2023 - Implementazione dell'osservabilità delle applicazioni
- AWS re:Invent 2021 - Automatizzazione intelligente delle operazioni cloud
- AWS re:Invent 2022 - Configurazione di controlli su larga scala nel proprio ambiente AWS
- AWS re:Invent 2022 - Automatizzazione della gestione e della conformità delle patch utilizzando AWS
- AWS re:Invent 2022 - In che modo Amazon utilizza metriche migliori per migliorare le prestazioni del sito Web
- AWS re:Invent 2023 - Prenditi una pausa: diagnostica e risolvi i problemi di prestazioni con Amazon RDS

- AWS re:Invent 2021 - {New Launch} Rileva e risolvi automaticamente i problemi con Amazon Guru DevOps
- AWS re:Invent 2023 - Centralizza le tue operazioni

Esempi correlati:

- CloudWatch Registri: personalizza gli allarmi

PERF05-BP06 Conserva il carico di lavoro e i servizi up-to-date

Resta up-to-date su nuovi servizi e funzionalità cloud per adottare funzionalità efficienti, rimuovere problemi e migliorare l'efficienza complessiva delle prestazioni del tuo carico di lavoro.

Anti-pattern comuni:

- Si ritiene che l'architettura corrente diventi statica e non venga aggiornata nel corso del tempo.
- Non si dispone di sistemi né si esegue regolarmente una valutazione per la compatibilità di software e pacchetti aggiornati con il carico di lavoro.

Vantaggi derivanti dall'adozione di questa best practice: stabilendo un processo per rimanere aggiornato up-to-date su nuovi servizi e offerte, puoi adottare nuove funzionalità e funzionalità, risolvere problemi e migliorare le prestazioni del carico di lavoro.

Livello di rischio associato se questa best practice non fosse adottata: basso

Guida all'implementazione

Valuta i modi per migliorare le prestazioni man mano che nuovi servizi, modelli di progettazione e funzionalità di prodotti diventano disponibili. Determina in che modo possono migliorare le prestazioni o aumentare l'efficienza del carico di lavoro tramite valutazioni, discussioni interne o analisi esterne. Definisci un processo per valutare gli aggiornamenti, le nuove funzionalità e i servizi pertinenti per il tuo carico di lavoro. Ad esempio, crea un proof of concept che utilizza le nuove tecnologie o consultati con un gruppo interno. Quando provi nuove idee o servizi, esegui test delle prestazioni per misurare l'impatto sulle prestazioni del carico di lavoro.

Passaggi dell'implementazione

- Esegui l'inventario del tuo carico di lavoro: esegui l'inventario di software e architettura del carico di lavoro e identifica i componenti da aggiornare.

- Identifica le origini dell'aggiornamento: identifica novità e origini dell'aggiornamento relative ai componenti del carico di lavoro. Ad esempio, puoi iscriverti al AWS blog What's New at per i prodotti che corrispondono al tuo componente di carico di lavoro. Puoi iscriverti al RSS feed o gestire le tue iscrizioni e-mail.
- Definisci un programma di aggiornamento: definisci un programma per valutare nuovi servizi e funzionalità per il tuo carico di lavoro.
  - Puoi utilizzare AWS Systems Manager Inventory per raccogliere i metadati del sistema operativo (OS), delle applicazioni e delle istanze dalle tue EC2 istanze Amazon e capire rapidamente quali istanze eseguono il software e le configurazioni richieste dalla tua politica software e quali istanze devono essere aggiornate.
- Valuta il nuovo aggiornamento: individua le modalità di aggiornamento dei componenti del carico di lavoro. Sfrutta l'agilità del cloud per testare in modo semplice e rapido il modo in cui le nuove funzionalità possono migliorare il carico di lavoro per ottenere efficienza delle prestazioni.
- Utilizza l'automazione: sfrutta l'automazione del processo di aggiornamento per ridurre il livello di impegno per implementare le nuove funzionalità e limitare gli errori causati dai processi manuali.
  - Puoi utilizzare CI/CD per aggiornare AMIs automaticamente le immagini dei container e altri elementi relativi alla tua applicazione cloud.
  - È possibile utilizzare strumenti come AWS Systems Manager Patch Manager per automatizzare il processo di aggiornamento del sistema e pianificare l'attività utilizzando le finestre di manutenzione di AWS Systems Manager.
- Documenta il processo: documenta il tuo processo di valutazione di aggiornamenti e nuovi servizi. Fornisci ai proprietari il tempo e lo spazio necessari per ricercare, testare, sperimentare e convalidare aggiornamenti e nuovi servizi. Fate riferimento ai requisiti aziendali documentati e aiutateci KPIs a stabilire le priorità degli aggiornamenti che avranno un impatto positivo sull'azienda.

Risorse

Documenti correlati:

- Blog AWS
- Cosa c'è di nuovo con AWS
- Implementazione di up-to-date immagini con pipeline automatizzate di EC2 Image Builder

Video correlati:

- AWS RE:InForce 2022 - Automatizzazione della gestione e della conformità delle patch utilizzando AWS
- All Things Patch: | Eventi AWS Systems ManagerAWS

Esempi correlati:

- Inventory and Patch Management
- One Observability Workshop

PERF05-BP07 Rivedi le metriche a intervalli regolari

Nell'ambito della manutenzione ordinaria o in risposta a eventi o incidenti, esamina i parametri raccolti. Stabilisci quali di questi parametri sono fondamentali per risolvere i problemi e quali altri parametri aggiuntivi, se monitorati, possono contribuire a identificare, affrontare o prevenire i problemi.

Anti-pattern comuni:

- Si lascia che i parametri rimangano in uno stato di allarme per un lungo periodo di tempo.
- Creazione di allarmi non utilizzabili da un sistema di automazione.

Vantaggi dell'adozione di questa best practice: esamina in modo continuo i parametri raccolti per verificare che identifichino, risolvano o prevengano adeguatamente i problemi. I parametri possono anche diventare obsoleti se lasciati in uno stato di allarme per un lungo periodo di tempo.

Livello di rischio associato se questa best practice non fosse adottata: medio

Guida all'implementazione

Migliora continuamente la raccolta e il monitoraggio dei parametri. Nell'ambito della risposta a incidenti ed eventi, valuta quali parametri sono stati utili per affrontare il problema e quali sarebbero stati utili ma non sono attualmente misurati. Questo metodo ti aiuterà a migliorare la qualità dei parametri raccolti, in modo da prevenire o risolvere in modo più rapido gli incidenti futuri.

Nell'ambito della risposta a incidenti ed eventi, valuta quali parametri sono stati utili per affrontare il problema e quali sarebbero stati utili ma non sono attualmente misurati. Queste considerazioni ti aiuteranno a migliorare la qualità dei parametri raccolti, così da prevenire o risolvere più rapidamente gli incidenti futuri.

Passaggi dell'implementazione

- Definisci metriche: stabilisci metriche in termini di prestazioni critiche da monitorare, allineate all'obiettivo del carico di lavoro, incluse metriche quali il tempo di risposta e l'utilizzo delle risorse.
- Stabilisci una base: imposta un valore di base e auspicabile per ciascuna metrica. La base e deve fornire i punti di riferimento per identificare deviazioni o anomalie.
- Imposta una cadenza: imposta una cadenza (ad esempio, settimanale o mensile) per rivedere le metriche più critiche.
- Identifica i problemi di prestazioni: durante ogni revisione, valuta tendenze e deviazione dai valori di base. Cerca eventuali colli di bottiglia o anomalie nelle prestazioni. Per i problemi identificati, esegui un'analisi approfondita delle cause principali per comprendere il motivo più importante alla base del problema.
- Individua le azioni correttive: utilizza l'analisi per identificare le azioni correttive, come l'ottimizzazione dei parametri, la correzione di bug e il dimensionamento delle risorse.
- Documenta gli esiti: documenta gli esiti, compresi i problemi identificati, le cause principali e le azioni correttive.
- Itera migliora: valuta e migliora continuamente il processo di revisione delle metriche. Usa le indicazioni apprese dalla revisione precedente per migliorare il processo nel tempo.

Risorse

Documenti correlati:

- CloudWatchDocumentazione
- Raccogli parametri e log da Amazon EC2 Instances e server locali con l'agente CloudWatch
- Interroga le tue metriche con Metrics Insights CloudWatch
- Partner per il monitoraggio, la registrazione e le prestazioni AWS Partner Network
- Documentazione di X-Ray

Video correlati:

- AWS re:Invent 2022 - Configurazione di controlli su larga scala nel proprio ambiente AWS
- AWS re:Invent 2022 - In che modo Amazon utilizza metriche migliori per migliorare le prestazioni del sito Web
- AWS re:Invent 2023 - Creazione di un'efficace strategia di osservabilità

- AWS Summit SF 2022 - Osservabilità completa e monitoraggio delle applicazioni con AWS
- AWS re:Invent 2023 - Prenditi una pausa: diagnostica e risolvi i problemi di prestazioni con Amazon RDS

Esempi correlati:

- Creazione di una dashboard con Amazon QuickSight
- CloudWatch Pannelli di controllo

Ottimizzazione dei costi

Il pilastro dell'ottimizzazione dei costi include la possibilità di eseguire sistemi per offrire valore aggiunto al prezzo più basso. Le linee guida con le prescrizioni sull'implementazione sono disponibili nel whitepaper sul pilastro dell'ottimizzazione dei costi.

Aree delle best practice

- Implementazione della gestione finanziaria del cloud
- Comprensione delle spese e dell'utilizzo
- Risorse convenienti in termini di costo
- Gestione delle risorse di domanda e offerta
- Ottimizzazione nel tempo

Implementazione della gestione finanziaria del cloud

Domanda

- COST1. Come implementi la gestione finanziaria nel cloud?

COST1. Come implementi la gestione finanziaria nel cloud?

L'implementazione di Cloud Financial Management aiuta le organizzazioni a realizzare valore aziendale e successo finanziario ottimizzando costi e utilizzo e aumentando la scalabilità AWS.

Best practice

- COST01-BP01 Stabilire la titolarità dell'ottimizzazione dei costi

- COST01-BP02 Stabilire una partnership tra finanza e tecnologia
- COST01-BP03 Stabilisci budget e previsioni per il cloud
- COST01-BP04 Implementare la consapevolezza dei costi nei processi organizzativi
- COST01-BP05 Segnala e notifica sull'ottimizzazione dei costi
- COST01-BP06 Monitora i costi in modo proattivo
- COST01-BP07 Resta aggiornato up-to-date sulle nuove release di servizio
- COST01-BP08 Creare una cultura attenta ai costi
- COST01-BP09 Quantifica il valore aziendale grazie all'ottimizzazione dei costi

COST01-BP01 Stabilire la titolarità dell'ottimizzazione dei costi

Crea un team (Cloud Business Office, Cloud Center of Excellence o FinOps team) responsabile della creazione e del mantenimento della consapevolezza dei costi in tutta l'organizzazione. Il responsabile dell'ottimizzazione dei costi può essere un individuo o un team (sono necessarie persone provenienti da team finanziari, tecnologici e aziendali) che ha una comprensione dell'intera organizzazione e degli aspetti finanziari legati al cloud.

Livello di rischio associato se questa best practice non fosse adottata: elevato

Guida all'implementazione

Si tratta dell'introduzione di una funzione o di un team di Cloud Business Office (CBOCCOE) o Cloud Center of Excellence () responsabile della creazione e del mantenimento di una cultura della consapevolezza dei costi nel cloud computing. Questa funzione può essere una figura professionale già in organico, un team all'interno della tua organizzazione o un nuovo team di parti interessate chiave dei settori finanza, tecnologia e organizzazione provenienti da tutta l'azienda.

La funzione (individuo o team) stabilisce le priorità e dedica la parte prevista del proprio tempo alle attività di gestione e ottimizzazione dei costi. In un'organizzazione di dimensioni ridotte, la quantità di tempo dedicata dalla funzione potrebbe essere inferiore rispetto a quella dedicata da una funzione a tempo pieno in un'azienda di dimensioni maggiori.

La funzione richiede un approccio multidisciplinare, con capacità di gestione dei progetti, data science, analisi finanziaria e sviluppo di software o infrastruttura. Può migliorare l'efficienza del carico di lavoro eseguendo ottimizzazioni dei costi all'interno di tre diversi tipi di responsabilità:

- Centralizzato: tramite team designati come FinOps team, team Cloud Financial Management (CFM), Cloud Business Office (CBO) o Cloud Center of Excellence (CCoE), i clienti possono

progettare e implementare meccanismi di governance e promuovere le migliori pratiche a livello aziendale.

- Team decentralizzati: influenzano i team tecnologici per ottimizzare i costi.
- Team ibridi: una combinazione di team centralizzati e decentralizzati può collaborare per eseguire l'ottimizzazione dei costi.

La funzione può essere valutata in base alla sua capacità di eseguire e conseguire risultati rispetto agli obiettivi di ottimizzazione dei costi (ad esempio in base a metriche di efficienza dei carichi di lavoro).

Un fattore chiave per il successo di questa funzione è la disponibilità di sponsorizzazione da parte del management. Lo sponsor deve essere un sostenitore del consumo efficiente del cloud e fornire alla funzione un supporto in caso di escalation, per garantire che le attività di ottimizzazione dei costi vengano trattate con il livello di priorità definito dall'organizzazione. In caso contrario, le linee guida possono essere ignorate e non verrà data priorità alle opportunità di riduzione dei costi. Insieme, lo sponsor e il team dell'organizzazione possono aiutare a utilizzare il cloud in modo efficiente e generare valore aziendale.

Se disponi del piano di supporto Business Enterprise-On-Ramp o Enterprise e hai bisogno di aiuto per creare questo team o questa funzione, contatta gli esperti di Cloud Financial Management (CFM) tramite il tuo account team.

Passaggi dell'implementazione

- Definizione dei membri chiave: tutte le parti rilevanti della tua organizzazione devono contribuire ed essere interessate alla gestione dei costi. I team comuni all'interno delle organizzazioni includono in genere: responsabili finanziari, proprietari delle applicazioni o dei prodotti, team di gestione e tecnici (DevOps). Alcuni soggetti sono impegnati a tempo pieno (ad esempio quelli di tipo finanziario o tecnico), mentre altri sono coinvolti periodicamente secondo necessità. Gli individui o i team che si esibiscono CFM necessitano delle seguenti competenze:
  - Sviluppo software: competenze inerenti allo sviluppo di software, in caso di sviluppo di script e funzioni di automazione.
  - Progettazione dell'infrastruttura: per implementare script, automatizzare processi e comprendere in che modo vengono allocati risorse e servizi.
  - Acume operativo: CFM consiste nell'operare sul cloud in modo efficiente misurando, monitorando, modificando, pianificando e scalando l'uso efficiente del cloud.

- Definizione di obiettivi e metriche: la funzione deve fornire valore all'organizzazione in modi diversi. Questi obiettivi sono definiti e si evolvono continuamente con l'evolversi dell'organizzazione. Tra le attività più comuni figurano la creazione e l'esecuzione di programmi di formazione sull'ottimizzazione dei costi in tutta l'organizzazione, lo sviluppo di standard a livello aziendale, come monitoraggio ed elaborazione di report per l'ottimizzazione dei costi, e la definizione degli obiettivi di ottimizzazione dei carichi di lavoro. Inoltre, è necessario comunicare regolarmente all'organizzazione la relativa capacità di ottimizzazione dei costi.

  È possibile definire indicatori chiave di prestazione basati sul valore o sul costo (). KPIs Quando si definisce ilKPIs, è possibile calcolare il costo previsto in termini di efficienza e risultati aziendali attesi. Basate sul valore, collegate KPIs le metriche di costo e utilizzo ai fattori di valore aziendale e aiutate a razionalizzare le variazioni di spesa. AWS Il primo passo per ricavare valori basati sul valore KPIs consiste nel lavorare insieme, a livello interorganizzativo, per selezionare e concordare un set standard di. KPIs
- Definizione di una cadenza regolare: il gruppo (team finanziario, tecnologico e aziendale) deve riunirsi regolarmente per rivedere le metriche e gli obiettivi. Una periodicità tipica implica la revisione dello stato dell'organizzazione, la revisione dei programmi attualmente in esecuzione e la revisione dei parametri finanziari e di ottimizzazione generali. Quindi, per i carichi di lavoro chiave, è opportuno elaborare report più dettagliati.

  Durante queste riunioni periodiche è possibile analizzare l'efficienza (costo) dei carichi di lavoro e il risultato aziendale. Ad esempio, un incremento del 20% dei costi di un carico di lavoro potrebbe essere determinato dall'aumento dell'utilizzo da parte dei clienti. In questo caso, l'incremento del 20% dei costi può essere interpretato come investimento. Queste chiamate a cadenza regolare possono aiutare i team a identificare i valori KPIs che danno significato all'intera organizzazione.

Risorse

Documenti correlati:

- Blog AWS CCOE
- Creating Cloud Business Office
- CCOE- Centro di eccellenza cloud

Video correlati:

- Storia di CCOE successo di Vanguard

Esempi correlati:

- Utilizzo di un Cloud Center of Excellence (CCOE) per trasformare l'intera azienda
- Costruire un CCOE programma per trasformare l'intera azienda
- 7 insidie da evitare durante la costruzione CCOE

COST01-BP02 Stabilire una partnership tra finanza e tecnologia

Coinvolgi i team finanziari e tecnologici nelle discussioni su costi e utilizzo in tutte le fasi del tuo percorso verso il cloud. I team si riuniscono regolarmente e discutono argomenti quali obiettivi e target organizzativi, stato attuale di costi e utilizzo e pratiche finanziarie e contabili.

Livello di rischio associato se questa best practice non fosse adottata: elevato

Guida all'implementazione

I team tecnologici possono innovare più rapidamente nel cloud grazie a cicli di approvazione, approvvigionamento e implementazione dell'infrastruttura più brevi. Può trattarsi di una novità per le organizzazioni finanziarie che in precedenza erano abituate a eseguire processi dispendiosi, in termini di tempo e di risorse, per acquistare e distribuire capitale in data center e on-premises, allocando i costi solo in fase di approvazione del progetto.

Dal punto di vista delle organizzazioni finanziarie e addette all'approvvigionamento, il processo di elaborazione del piano degli investimenti, della richiesta, dell'approvazione e dell'approvvigionamento degli investimenti e dell'installazione dell'infrastruttura fisica è stato interiorizzato e standardizzato da decenni:

- I team di progettazione o IT sono in genere i richiedenti
- I vari team finanziari fungono da approvatori e addetti all'approvvigionamento
- I team operativi raccolgono, impilano e consegnano l'infrastruttura ready-to-use

Con l'adozione del cloud, l'approvvigionamento e il consumo dell'infrastruttura non sono più vincolati da una catena di dipendenze. Nel modello cloud, i team tecnologici e del prodotto non sono più semplici sviluppatori, ma anche operatori e proprietari dei loro prodotti, responsabili della maggior parte delle attività storicamente associate ai team finanziari e operativi, compresi l'approvvigionamento e l'implementazione.

Quanto in realtà è necessario per il provisioning delle risorse cloud è un account e il set appropriato di autorizzazioni, Questo è anche ciò che riduce i rischi IT e finanziari; il che significa che i team sono sempre a pochi clic o API chiamate dal terminare le risorse cloud inattive o non necessarie. Ciò inoltre consente ai team tecnologici di velocizzare l'innovazione, grazie all'agilità e alla capacità di potenziare e quindi ridimensionare i vari progetti sperimentali. Se da un lato la natura variabile del

consumo del cloud può influenzare la prevedibilità dal punto di vista del processo di elaborazione del piano degli investimenti e delle previsioni, il cloud fornisce alle organizzazioni la capacità di ridurre il costo del provisioning eccessivo e contemporaneamente il costo delle opportunità associato a un provisioning insufficiente di carattere conservativo.

Stabilisci una collaborazione tra le principali parti interessate finanziarie e tecnologiche per creare una conoscenza condivisa degli obiettivi organizzativi e sviluppare meccanismi che consentano il successo finanziario nel modello di spesa variabile del cloud computing. I team pertinenti all'interno della tua organizzazione devono essere coinvolti nelle discussioni su costi e utilizzo in tutte le fasi del tuo percorso verso il cloud; tra di essi vi sono:

- Responsabili finanziari: i controllori finanziariCFOs, i pianificatori finanziari, gli analisti aziendali, gli acquisti, l'approvvigionamento e la contabilità fornitori devono comprendere il modello cloud di consumo, le opzioni di acquisto e il processo di fatturazione mensile. I team finanziari devono collaborare con i team tecnologici per creare e divulgare a livello aziendale una narrazione del

valore IT che aiuti i team aziendali a comprendere lo stretto legame tra spesa in tecnologie e risultati aziendali. In questo modo, la spesa tecnologica viene considerata non tanto come un costo, quanto piuttosto come un vero e proprio investimento. A causa delle differenze fondamentali tra il cloud (ad esempio il tasso di variazione dell'utilizzo, il pagamento in base al consumo o a scaglioni, i modelli di prezzo e le informazioni dettagliate su fatturazione e utilizzo) e le operazioni on-premises, è essenziale che l'organizzazione finanziaria capisca in che modo l'utilizzo del cloud può influire sugli aspetti aziendali, tra cui processi di approvvigionamento, monitoraggio degli incentivi, allocazione dei costi e bilanci.

- Responsabili tecnologici: i responsabili tecnologici (inclusi i proprietari di prodotti e applicazioni) devono essere a conoscenza dei requisiti finanziari (ad esempio i vincoli di budget) e dei requisiti aziendali (ad esempio i contratti sul livello di servizio). In questo modo, il carico di lavoro può essere implementato in modo opportuno per raggiungere gli obiettivi desiderati dall'azienda.

La collaborazione tra finanza e tecnologia offre i seguenti vantaggi:

- I team finanziari e tecnologici hanno una visibilità quasi in tempo reale su costi e utilizzo.
- I team finanziari e tecnologici stabiliscono una procedura operativa standard per gestire le variazioni di spesa nel cloud.
- Gli stakeholder finanziari agiscono in qualità di consulenti strategici per quanto riguarda il modo in cui il capitale viene utilizzato per acquistare sconti sugli impegni (ad esempio, Reserved Instances o AWS Savings Plans) e il modo in cui il cloud viene utilizzato per far crescere l'organizzazione.
- I processi di approvvigionamento e di contabilità esistenti vengono applicati al cloud.
- I team finanziari e tecnologici collaborano alla previsione AWS dei costi e dell'utilizzo futuri per allineare e costruire i budget organizzativi.
- La comunicazione all'interno dell'organizzazione migliora attraverso un linguaggio condiviso e una comprensione comune dei concetti finanziari.

Altre parti interessate all'interno della tua organizzazione che devono essere coinvolti nelle discussioni su costi e utilizzo includono:

- Proprietari delle business unit: i proprietari delle business unit devono comprendere il modello aziendale del cloud in modo da indirizzare l'operato delle business unit e di tutta l'azienda. Questa conoscenza del cloud è fondamentale quando è necessario prevedere la crescita e l'utilizzo del carico di lavoro, ma anche quando si valutano le diverse opzioni di acquisto, come le istanze riservate o i Savings Plans.

- Team di progettazione: stabilire una partnership tra i team finanziari e tecnologici è essenziale per creare una cultura attenta ai costi che incoraggi gli ingegneri ad agire sulla gestione finanziaria del cloud (). CFM Uno dei problemi più comuni dei nostri professionisti delle CFM operazioni finanziarie e dei team finanziari è far sì che gli ingegneri comprendano l'intero business sul cloud, seguano le migliori pratiche e intraprendano le azioni consigliate.
- Terze parti: se la vostra organizzazione utilizza terze parti (ad esempio consulenti o strumenti), assicuratevi che siano in linea con i vostri obiettivi finanziari e che possano dimostrare sia l'allineamento attraverso i loro modelli di coinvolgimento sia un ritorno sull'investimento (). ROI In genere, le terze parti contribuiscono alla creazione di report e all'analisi di eventuali carichi di lavoro da esse gestiti, e forniscono anche l'analisi dei costi relativi ai carichi di lavoro da esse progettati.

L'implementazione CFM e il raggiungimento del successo richiedono la collaborazione tra i team finanziari, tecnologici e aziendali e un cambiamento nel modo in cui la spesa per il cloud viene comunicata e valutata all'interno dell'organizzazione. Includi i team di progettazione in modo da renderli partecipi delle discussioni su costi e utilizzi in tutte le fasi e incoraggiali ad attenersi alle best practice e ad adottare le azioni concordate.

Passaggi dell'implementazione

- Definizione dei membri chiave: verifica che tutti i membri rilevanti dei team finanziari e tecnologici partecipino alla partnership. I membri del team finanziario interessati saranno quelli che hanno a che fare con la fatturazione dei servizi cloud. Si tratta in genere di controllori finanziariCFOs, pianificatori finanziari, analisti aziendali, addetti all'approvvigionamento e all'approvvigionamento. I membri tecnologici sono in genere i proprietari di prodotti e applicazioni, manager tecnici e rappresentanti di tutti i team che si basano sul cloud. Altri membri possono includere i responsabili di business unit, ad esempio il marketing che influenzerà l'utilizzo dei prodotti, e terze parti, come i consulenti, necessari per garantire l'allineamento agli obiettivi e meccanismi e per fornire assistenza nell'elaborazione dei report.
- Definizione degli argomenti di discussione: definisci gli argomenti comuni tra i team o che necessitano di una comprensione condivisa. Segui il costo dal momento in cui viene creato, fino al pagamento della fattura. Prendi nota di tutti i membri coinvolti e dei processi organizzativi che devono essere applicati. Comprendi ogni fase o processo e le informazioni associate, come i modelli di prezzo disponibili, i prezzi a scaglioni, i modelli di sconto, il budget e i requisiti finanziari.
- Definizione di una regolare cadenza: per creare una partnership tra team finanziari e tecnologici, definisci la periodicità delle comunicazioni per creare e gestire l'allineamento. Il gruppo deve riunirsi regolarmente in base ai propri obiettivi e parametri. Una periodicità tipica implica la revisione dello

stato dell'organizzazione, la revisione dei programmi attualmente in esecuzione e la revisione dei parametri finanziari e di ottimizzazione generali. Quindi, per i carichi di lavoro chiave, è opportuno elaborare report più dettagliati.

Risorse

Documenti correlati:

- AWS Blog di notizie

COST01-BP03 Stabilisci budget e previsioni per il cloud

Adatta i processi di previsione e di budgeting organizzativi esistenti in modo che siano compatibili con la natura altamente variabile dei costi e dell'utilizzo del cloud. I processi devono essere dinamici, utilizzando algoritmi basati su tendenze o fattori chiave aziendali o una combinazione di entrambi.

Livello di rischio associato se questa best practice non fosse adottata: elevato

Guida all'implementazione

Nelle tradizionali configurazioni IT on-premises, i clienti spesso devono affrontare la sfida di pianificare i costi fissi che variano solo occasionalmente, di solito con i nuovi acquisti di hardware e servizi IT per soddisfare i picchi di domanda. Al contrario, Cloud AWS adotta un approccio diverso, in cui i clienti pagano per le risorse che utilizzano in base alle loro effettive esigenze IT e aziendali. Nell'ambiente cloud, la domanda può variare su base mensile, giornaliera o persino oraria.

Il cloud offre efficienza, velocità e agilità, consolidando un modello di costo e utilizzo altamente variabile. I costi possono diminuire o talvolta aumentare in seguito all'incremento dell'efficienza dei carichi di lavoro o all'implementazione di nuovi carichi di lavoro e funzionalità. Man mano che i carichi di lavoro scalano per servire la clientela in crescita, l'utilizzo e i costi del cloud aumentano di conseguenza a causa del maggiore uso di risorse. Questa flessibilità dei servizi cloud si estende ai costi e alle previsioni, offrendo un certo grado di elasticità.

Per ottenere la pianificazione più accurata possibile, è essenziale allinearsi prontamente a queste mutevoli esigenze aziendali e ai fattori trainanti della domanda. I tradizionali processi di budget dell'organizzazione devono cambiare per far fronte a questa variabilità.

Valuta la modellazione dei costi mentre prevedi la spesa dei nuovi carichi di lavoro. La modellazione dei costi crea una comprensione di base dei costi previsti del cloud, che consente di eseguire il costo

totale di proprietà (TCO), il ritorno sull'investimento (ROI) e altre analisi finanziarie, stabilire obiettivi e aspettative con le parti interessate e identificare opportunità di ottimizzazione dei costi.

È necessario che l'organizzazione comprenda la definizione dei costi e i raggruppamenti accettati. Il livello di dettaglio usato per le previsioni può variare in base alla struttura dell'organizzazione e ai flussi di lavoro interni. Scegli la granularità adatta ai tuoi requisiti specifici e alla configurazione dell'organizzazione. È importante comprendere a quale livello viene eseguita la previsione:

- Account di gestione o livello AWS Organizations : l'account di gestione è quello utilizzato per creare AWS Organizations. Le organizzazioni dispongono di un account di gestione in modo predefinito.
- Account collegato o membro: un account in Organizations è uno standard Account AWS che contiene AWS le tue risorse e le identità che possono accedere a tali risorse.
- Ambiente: un ambiente è una raccolta di AWS risorse che esegue una versione dell'applicazione. È possibile creare un ambiente con più account collegati o membri.
- Progetto: per progetto si intende una combinazione di obiettivi o attività prestabiliti da realizzare entro un determinato periodo di tempo È importante considerare il ciclo di vita del progetto durante la previsione.
- AWS servizi: gruppi o categorie come servizi di elaborazione o archiviazione in cui è possibile raggruppare AWS i servizi per le previsioni.
- Raggruppamento personalizzato: puoi creare gruppi personalizzati in base alle esigenze dell'organizzazione, ad esempio business unit, centri di costo, team, tag di allocazione dei costi, categorie di costi, account collegati o una combinazione di questi.

Individua i fattori aziendali che possono influire sui costi di utilizzo e fai le previsioni per ciascuno di essi separatamente per calcolare in anticipo l'utilizzo previsto. Alcuni fattori possono essere collegati ai team IT e di prodotto dell'organizzazione. Altri fattori aziendali, come eventi di marketing, promozioni, espansioni geografiche, fusioni e acquisizioni, sono noti ai responsabili dell'area vendite, marketing e commerciale, quindi è importante collaborare e tenere conto anche di tutti questi fattori trainanti della domanda.

È possibile utilizzarlo AWS Cost Explorerper la previsione basata sulle tendenze in un intervallo di tempo futuro definito in base alle spese passate. AWS Cost Explorer segmenta i dati storici in base ai tipi di addebito (ad esempio, istanze riservate) e utilizza una combinazione di apprendimento automatico e modelli basati su regole per prevedere individualmente la spesa per tutti i tipi di addebito.

Dopo aver stabilito il processo di previsione e creato i modelli, puoi Budget AWSutilizzarlo per impostare budget personalizzati a livello granulare specificando il periodo di tempo, la ricorrenza o l'importo (fisso o variabile) e aggiungere filtri come servizio e tag. Regione AWS Il budget è generalmente definito per un solo anno e rimane fisso, richiedendo il rispetto rigoroso di tutte le parti coinvolte. Al contrario, le previsioni sono più flessibili, consentono adattamenti nel corso dell'anno e forniscono proiezioni dinamiche su un periodo di uno, due o tre anni. I budget e le previsioni svolgono un ruolo determinante nella definizione delle aspettative finanziarie tra le varie parti interessate tecnologiche e aziendali. Una previsione e un'implementazione accurate rendono responsabili anche le parti interessate che sono direttamente coinvolte nella gestione dei costi di provisioning e possono aumentare la loro consapevolezza generale dei costi.

Per essere informati sulle prestazioni dei budget esistenti, puoi creare e pianificare report Budget AWS da inviare tramite e-mail alle parti interessate con cadenza regolare. Puoi anche creare avvisi di Budget AWS basati sui costi effettivi, ovvero avvisi intrinsecamente reattivi, oppure sui costi previsti, ossia avvisi che consentono di implementare tempestivamente azioni correttive a fronte di potenziali eventi di superamento dei costi. Puoi ricevere un avviso quando il costo o l'utilizzo supera un determinato livello oppure si prevede che superi l'importo definito nel budget.

Modifica i processi di budget e previsione esistenti per renderli più dinamici utilizzando gli algoritmi basati sulle tendenze con i costi storici come input e gli algoritmi basati sui fattori aziendali, ad esempio il lancio di nuovi prodotti, l'espansione regionale o i nuovi ambienti per i carichi di lavoro, ideali per un ambiente di spesa dinamico e variabile. Dopo aver determinato la previsione basata sulle tendenze utilizzando Cost Explorer o qualsiasi altro strumento, utilizzala AWS Pricing Calculatorper stimare il caso AWS d'uso e i costi futuri in base all'utilizzo previsto (traffico o EC2 istanze Amazon richieste). requests-per-second

Controlla l'accuratezza di questa previsione perché i budget devono essere impostati sulla base di questi calcoli e queste stime. Monitora la precisione e l'efficacia delle previsioni dei costi del cloud integrate. Esamina con regolarità la spesa effettiva rispetto alla tua previsione e apporta le modifiche necessarie per ottenere una maggiore accuratezza. Controlla la varianza prevista ed esegui l'analisi della causa principale della varianza indicata per intervenire e modificare le previsioni.

Come indicato in COST01-BP02 Stabilire una partnership tra finanza e tecnologia, è importante promuovere partnership e cadenza tra IT, finanza e altre parti interessate per verificare che tutti utilizzino gli stessi strumenti o processi per garantire la coerenza. Nei casi in cui si rendano necessarie modifiche del budget, l'incremento della frequenza delle occasioni di contatto permette di intervenire e reagire più tempestivamente.

Passaggi dell'implementazione

- Definisci il linguaggio dei costi all'interno dell'organizzazione: crea un linguaggio di AWS costo comune all'interno dell'organizzazione con più dimensioni e raggruppamenti. Assicurati che le parti interessate comprendano la granularità delle previsioni, i modelli di prezzo e il livello delle previsioni dei costi.
- Analizza le previsioni basate sulle tendenze: utilizza strumenti per le previsioni basate sulle tendenze, come AWS Cost Explorer e Amazon Forecast. Analizza i costi di utilizzo rispetto a più dimensioni, come servizi, account, tag e categorie di costi. Se sono necessarie previsioni avanzate, importa i dati relativi a AWS costi e utilizzo (CUR) in Amazon Forecast (che applica la regressione lineare come forma di apprendimento automatico per le previsioni).
- Analizza le previsioni basate sui fattori aziendali: identifica l'impatto dei fattori aziendali sull'utilizzo del cloud e fai previsioni per ciascuno di essi separatamente per calcolare in anticipo il costo di utilizzo previsto. Collabora a stretto contatto con i responsabili delle business unit e le parti interessate per comprendere l'impatto dei nuovi fattori aziendali e calcolare le variazioni dei costi previste per definire budget accurati.
- Aggiorna i processi di previsione e budget esistenti: definisci i tuoi processi di previsione del budget in base ai metodi di previsione adottati, ad esempio basati sulle tendenze, basati sui fattori di aziendali o su una combinazione di entrambi i metodi di previsione. I budget devono essere calcolati, realistici e basati sulle previsioni.
- Configura avvisi e notifiche: utilizza gli Budget AWS avvisi e il rilevamento delle anomalie dei costi per ricevere avvisi e notifiche.
- Esegui revisioni periodiche con le principali parti interessate: ad esempio, è necessario allinearsi ai cambiamenti nella direzione dell'azienda e nell'utilizzo con le parti interessate dell'IT, della finanza, dei team della piattaforma e di altre aree dell'azienda.

Risorse

Documenti correlati:

- AWS Cost Explorer
- AWS Cost and Usage Report
- Forecasting with Cost Explorer
- QuickSight Previsioni Amazon
- Amazon Forecast

- Budget AWS

Video correlati:

- Come posso utilizzarlo Budget AWS per tenere traccia delle mie spese e del mio utilizzo
- AWS Serie di ottimizzazione dei costi: Budget AWS

Esempi correlati:

- Understand and build driver-based forecasting
- How to establish and drive a forecasting culture
- How to improve your cloud cost forecasting
- Using the right tools for your cloud cost forecasting

COST01-BP04 Implementare la consapevolezza dei costi nei processi organizzativi

Implementa la consapevolezza dei costi e crea trasparenza e funzionalità di controllo in processi nuovi o esistenti che influiscono sull'utilizzo e sfrutta i processi esistenti per favorire la consapevolezza dei costi. Implementa la consapevolezza dei costi nella formazione dei dipendenti.

Livello di rischio associato se questa best practice non fosse adottata: elevato

Guida all'implementazione

La consapevolezza dei costi deve essere implementata nei processi organizzativi nuovi ed esistenti. Si tratta di un prerequisito fondamentale per altre best practice. È consigliabile riutilizzare e modificare i processi esistenti, laddove possibile, riducendo al minimo l'impatto sull'agilità e sulla velocità. Segnalate i costi del cloud ai team tecnologici e ai responsabili decisionali dei team aziendali e finanziari per aumentare la consapevolezza dei costi e stabilire indicatori chiave di efficienza (KPIs) per gli stakeholder finanziari e aziendali. Le seguenti raccomandazioni aiuteranno a implementare la consapevolezza dei costi nel carico di lavoro:

- Verifica che la gestione delle modifiche includa una misurazione dei costi per quantificare l'impatto finanziario delle modifiche. Questo aiuta a risolvere in modo proattivo le problematiche relative ai costi nonché a evidenziare i risparmi ottenuti.

- Verifica che l'ottimizzazione dei costi sia un componente fondamentale delle tue capacità operative. Ad esempio, puoi sfruttare gli attuali processi di gestione degli incidenti per analizzare e identificare la causa principale di anomalie di costi e utilizzo o delle eccedenze di costo.
- Accelera la riduzione dei costi e la realizzazione del valore aggiunto attraverso l'automazione o l'utilizzo di strumenti. Quando pensi ai costi di implementazione, inquadra la conversazione in modo da includere una componente relativa al ritorno sull'investimento (ROI) per giustificare l'investimento di tempo o denaro.
- Assegna i costi del cloud mediante l'implementazione delle policy di showback/chargeback per la spesa cloud, compresa la spesa per opzioni di acquisto basate su impegno, servizi condivisi e acquisti su marketplace, a supporto di un consumo del cloud maggiormente consapevole dei costi.
- Estendi i programmi di formazione e sviluppo esistenti per includere la formazione sulla consapevolezza dei costi in tutta l'organizzazione, comprese attività di formazione continua e certificazione. In questo modo, creerai un'organizzazione in grado di gestire in modo autonomo i costi e l'utilizzo.
- Sfrutta gli strumenti AWS nativi gratuiti come AWS Cost Anomaly DetectionBudget AWS, e Budget AWS Reports.

Quando le organizzazioni adottano costantemente le pratiche di Cloud Financial Management (CFM), tali comportamenti diventano radicati nel modo di lavorare e nel processo decisionale. Il risultato è una cultura più attenta ai costi, dagli sviluppatori che progettano una nuova born-in-the-cloud applicazione ai responsabili finanziari che analizzano questi nuovi investimenti nel cloud. ROI

Passaggi dell'implementazione

- Identificazione dei processi organizzativi pertinenti: ciascuna unità organizzativa esamina i propri processi e identifica quelli che influiscono su costi e utilizzo. Tutti i processi che determinano la creazione o la cessazione di una risorsa devono essere inclusi nella revisione. Individua i processi che possono supportare la consapevolezza dei costi nella tua azienda, ad esempio la gestione degli incidenti e la formazione.
- Stabilisci una cultura dell'autosufficienza e attenta ai costi: assicurati che tutte le parti interessate si allineino cause-of-change e che influiscano sui costi, in modo che comprendano i costi del cloud. Ciò consentirà all'organizzazione di definire una cultura consapevole dei costi autosufficiente finalizzata all'innovazione.
- Aggiornamento dei processi con la consapevolezza dei costi: la modifica dei processi avviene per renderli consapevoli dei costi. Il processo potrebbe richiedere ulteriori controlli preliminari, ad esempio la valutazione dell'impatto dei costi, oppure controlli successivi che attestino il verificarsi

dei cambiamenti previsti in termini di costi e utilizzo. I processi di supporto come la formazione e la gestione degli incidenti possono essere estesi per includere elementi relativi a costi e utilizzo.

Per ricevere assistenza, contatta CFM gli esperti tramite il team del tuo account oppure esplora le risorse e i documenti correlati riportati di seguito.

Risorse

Documenti correlati:

- AWS Gestione finanziaria nel cloud

Esempi correlati:

- Strategy for Efficient Cloud Cost Management
- Cost Control Blog Series #3: How to Handle Cost Shock
- Una guida per principianti a AWS Cost Management

COST01-BP05 Segnala e notifica sull'ottimizzazione dei costi

Imposta i budget per il cloud e configura i meccanismi per rilevare anomalie nell'utilizzo. Configura gli strumenti correlati per ricevere avvisi su costi e utilizzo rispetto a obiettivi predefiniti e ricevi notifiche quando l'utilizzo supera tali obiettivi. Organizza riunioni regolari per analizzare l'economicità dei tuoi carichi di lavoro e promuovere la consapevolezza dei costi.

Livello di rischio associato se questa best practice non fosse adottata: basso

Guida all'implementazione

È necessario inviare report regolari sull'ottimizzazione dei costi e sull'utilizzo all'interno dell'organizzazione. Puoi implementare sessioni dedicate per discutere le prestazioni in termini di costi o includere l'ottimizzazione dei costi nei regolari cicli di rendicontazione operativi per i tuoi carichi di lavoro. Utilizza servizi e strumenti per monitorare regolarmente le prestazioni in termini di costi e implementare opportunità di risparmio sui costi.

Visualizza i costi e l'utilizzo con più filtri e granularità utilizzando AWS Cost Explorer, che fornisce pannelli di controllo e report come i costi per servizio o per account, i costi giornalieri o i costi del marketplace. Monitora l'avanzamento di costi e utilizzo rispetto ai budget configurati attraverso i report di Budget AWS

Budget AWSUtilizzalo per impostare budget personalizzati per tenere traccia dei costi e dell'utilizzo e rispondere rapidamente agli avvisi ricevuti tramite e-mail o notifiche di Amazon Simple Notification Service SNS (Amazon) se superi la soglia. Imposta il periodo di budget preferito su giornaliero, mensile, trimestrale o annuale e crea limiti di budget specifici per rimanere informato sull'avanzamento dei costi e dell'utilizzo effettivi o previsti verso la soglia di budget. Puoi anche configurare avvisi e azioni da eseguire automaticamente o in base a un processo di approvazione a fronte di tali avvisi quando viene superato l'obiettivo del budget.

Implementa notifiche su costi e utilizzo per garantire che le variazioni di costi e utilizzo possano essere affrontate rapidamente in caso di imprevisti. AWS Cost Anomaly Detectionconsente di ridurre le sorprese in termini di costi e migliorare il controllo senza rallentare l'innovazione. AWS Cost Anomaly Detection identifica le spese anomale e le cause principali, il che aiuta a ridurre il rischio di sorprese nella fatturazione. Grazie a tre semplici passaggi, è possibile creare una funzione di controllo contestualizzata personalizzata e ricevere avvisi quando viene rilevata una spesa anomala.

Puoi anche utilizzare Amazon QuickSight con AWS Cost and Usage Report (CUR) dati, per fornire report altamente personalizzati con dati più granulari. Amazon ti QuickSight consente di pianificare report e ricevere e-mail periodiche di report sui costi e sull'utilizzo cronologici o sulle opportunità di risparmio sui costi. Dai un'occhiata alla nostra soluzione Cost Intelligence Dashboard (CID) costruita su Amazon QuickSight, che ti offre una visibilità avanzata.

Use AWS Trusted Advisor, che fornisce indicazioni per verificare se le risorse assegnate sono in linea con le AWS migliori pratiche per l'ottimizzazione dei costi.

Controlla le tue raccomandazioni Savings Plans tramite grafici visivi confrontandoli con i costi e l'utilizzo granulari. I grafici orari mostrano la spesa on demand insieme all'impegno verso i Savings Plans raccomandati, fornendo informazioni sui risparmi stimati, sulla copertura dei Savings Plans e sull'utilizzo dei Savings Plans. Questo aiuta le organizzazioni a capire in che modo i loro Savings Plans si applicano a ogni ora di spesa senza dover investire tempo e risorse nella creazione di modelli per analizzare la spesa stessa.

Crea periodicamente report contenenti informazioni dettagliate sui consigli di Savings Plans, Reserved Instances e Amazon AWS Cost Explorer per iniziare a ridurre EC2 i costi associati ai carichi di lavoro stazionari, alle risorse inattive e al sottoutilizzo. Individua e ammortizza la spesa associata all'utilizzo non ottimale del cloud relativamente alle risorse implementate. Con utilizzo non ottimale del cloud si intende la creazione di risorse dimensioni errate oppure la presenza di modelli di utilizzo del cloud diversi da quanto previsto. Segui le AWS best practice per ridurre gli sprechi o chiedi al tuo account team e al tuo partner di aiutarti a ottimizzare e ridurre i costi del cloud.

Genera regolarmente report per migliorare le opzioni di acquisto delle risorse al fine di ridurre il costo unitario dei carichi di lavoro. Le opzioni di acquisto come Savings Plans, Reserved Instances o Amazon EC2 Spot Instances offrono il massimo risparmio sui costi per carichi di lavoro con tolleranza ai guasti e consentono alle parti interessate (titolari di aziende, team finanziari e tecnici) di partecipare a queste discussioni sull'impegno.

Condividi i report che contengono opportunità o annunci di nuove release che possono aiutarti a ridurre il costo totale di proprietà () del cloud. TCO Adotta nuovi servizi, regioni, funzionalità, soluzioni o nuovi modi per migliorare ulteriormente la riduzione dei costi.

Passaggi dell'implementazione

- Configura Budget AWS: configura Budget AWS il tuo carico di lavoro su tutti gli account. Imposta un budget per la spesa complessiva dell'account e un budget per il carico di lavoro utilizzando i tag.
  - Well-Architected Labs: utilizzo di costi e governance
- Crea report sull'ottimizzazione dei costi: configura un ciclo regolare per discutere e analizzare l'efficienza del carico di lavoro. Utilizzando i parametri stabiliti, segnala i parametri raggiunti e il costo sostenuto per ottenerli. Identifica e correggi eventuali tendenze negative e individua tendenze positive che puoi favorire in tutta l'organizzazione. La rendicontazione dovrebbe coinvolgere i rappresentanti dei team e dei responsabili delle applicazioni, dei responsabili finanziari e dei principali responsabili delle decisioni in merito alla spesa per il cloud.

Risorse

Documenti correlati:

- AWS Cost Explorer
- AWS Trusted Advisor
- Budget AWS
- AWS Cost and Usage Report
- Budget AWS Best practice
- Amazon S3 Analytics

Esempi correlati:

- Well-Architected Labs: utilizzo di costi e governance

- Modi principali per iniziare a ottimizzare i costi del cloud AWS

COST01-BP06 Monitora i costi in modo proattivo

Implementa strumenti e pannelli di controllo per monitorare i costi in modo proattivo per il carico di lavoro. Rivedi regolarmente i costi utilizzando strumenti configurati o pronti all'uso e non limitarti a guardare solo i costi e le categorie quando ricevi le notifiche. Il monitoraggio e l'analisi dei costi aiutano in modo proattivo a individuare i trend positivi e a promuoverli nell'organizzazione.

Livello di rischio associato se questa best practice non fosse adottata: medio

Guida all'implementazione

Si consiglia di monitorare i costi e l'utilizzo all'interno dell'organizzazione in modo proattivo, e non solo in caso di eccezioni o anomalie. I pannelli di controllo con un'elevata visibilità in tutto l'ufficio o l'ambiente di lavoro garantiscono che le persone chiave abbiano accesso alle informazioni di cui hanno bisogno e dimostrano l'attenzione che l'organizzazione presta all'ottimizzazione dei costi. I pannelli di controllo visibili consentono di promuovere attivamente i risultati positivi e di implementarli in tutta l'organizzazione.

Crea una routine quotidiana o frequente da utilizzare AWS Cost Explorero qualsiasi altra dashboard come Amazon QuickSight per visualizzare i costi e analizzarli in modo proattivo. Analizza l'utilizzo e i costi del AWS servizio a AWS livello di account, di carico di lavoro o di livello di AWS servizio specifico con raggruppamento e filtraggio e verifica se sono previsti o meno. Utilizza tag e granularità a livello orario o di risorsa per filtrare e individuare i costi ricorrenti relativi alle risorse di maggiore utilizzo. Puoi anche creare report personalizzati con Cost Intelligence Dashboard, una QuickSight soluzione Amazon creata da AWS Solutions Architects, e confrontare i tuoi budget con i costi e l'utilizzo effettivi.

Passaggi dell'implementazione

- Crea report sull'ottimizzazione dei costi: configura un ciclo regolare per discutere e analizzare l'efficienza del carico di lavoro. Utilizzando i parametri stabiliti, segnala i parametri raggiunti e il costo sostenuto per ottenerli. Identifica e correggi eventuali tendenze negative e identifica le tendenze positive che puoi favorire in tutta l'organizzazione. L'elaborazione dei report deve coinvolgere i rappresentanti dei team applicativi e dei proprietari, dei team finanziari e di gestione.
- Crea e attiva una granularità giornaliera Budget AWSper i costi e l'utilizzo per intraprendere azioni tempestive per prevenire potenziali sforamenti dei costi: ti Budget AWS consente di configurare

le notifiche di avviso, in modo da rimanere informato se uno qualsiasi dei tuoi tipi di budget non supera le soglie preconfigurate. Il modo migliore per sfruttarlo Budget AWS è impostare i costi e l'utilizzo previsti come limiti, in modo che qualsiasi importo superiore al budget possa essere considerato una spesa eccessiva.

- Create AWS Cost Anomaly Detection for cost monitor: AWS Cost Anomaly Detectionutilizza una tecnologia avanzata di Machine Learning per identificare le spese anomale e le cause principali, in modo da poter intervenire rapidamente. Ti consente di configurare funzionalità di monitoraggio dei costi che definiscono i segmenti di spesa da valutare, ad esempio singoli servizi AWS , account membro, tag di allocazione dei costi e categorie di costo, nonché di impostare quando, dove e come riceverai le notifiche di avviso. Per ciascuna funzionalità di monitoraggio, puoi associare più sottoscrizioni agli avvisi per proprietari di azienda e team tecnologici, inclusi un nome, una soglia relativa all'impatto dei costi e la frequenza di avviso (avvisi singoli, riepilogo giornaliero, riepilogo settimanale) per ciascuna sottoscrizione.
- Usa AWS Cost Explorer o integra i tuoi AWS Cost and Usage Report (CUR) dati con QuickSight le dashboard di Amazon per visualizzare i costi della tua organizzazione: AWS Cost Explorer ha un' easy-to-useinterfaccia che ti consente di visualizzare, comprendere e gestire AWS i costi e l'utilizzo nel tempo. Cost Intelligence Dashboard è personalizzabile e accessibile e consente di creare le basi di uno strumento di gestione e ottimizzazione dei costi personalizzato.

Risorse

Documenti correlati:

- Budget AWS
- AWS Cost Explorer
- Daily Cost and Usage Budgets
- AWS Cost Anomaly Detection

Esempi correlati:

- Well-Architected Labs: visualizzazione
- Well-Architected Labs: visualizzazione avanzata
- Well-Architected Labs: Cloud Intelligence Dashboards
- Well-Architected Labs: visualizzazione dei costi
- AWS Cost Anomaly Detection Avvisi con Slack

COST01-BP07 Resta aggiornato up-to-date sulle nuove release di servizio

Consulta regolarmente esperti o AWS partner per valutare quali servizi e funzionalità offrono costi inferiori. AWS Consulta blog e altre fonti di informazione.

Livello di rischio associato se questa best practice non fosse adottata: medio

Guida all'implementazione

AWS aggiunge costantemente nuove funzionalità in modo da poter sfruttare le tecnologie più recenti per sperimentare e innovare più rapidamente. Potresti essere in grado di implementare nuovi AWS servizi e funzionalità per aumentare l'efficienza dei costi del tuo carico di lavoro. Consulta regolarmente la pagina sulla gestione dei costi AWS, il blog delle novitàAWS, il blog sulla gestione dei costi AWS, e Novità di AWS per informazioni su nuovi servizi e lanci di funzionalità. I post sulle novità forniscono una breve panoramica di tutti gli annunci relativi a AWS servizi, funzionalità e aree geografiche non appena vengono pubblicati.

Passaggi dell'implementazione

- Iscriviti ai blog: vai alle pagine dei AWS blog e iscriviti al blog What's New e ad altri blog pertinenti. Puoi registrarti nella pagina delle preferenze di comunicazione con il tuo indirizzo e-mail.
- Iscriviti alle AWS notizie: consulta regolarmente il AWS News Blog e What's New with AWS per informazioni sulle nuove versioni di servizi e funzionalità. Iscriviti al RSS feed o con la tua email per seguire gli annunci e i comunicati.
- Segui le riduzioni AWS dei prezzi: le riduzioni regolari dei prezzi di tutti i nostri servizi sono state un modo standard per trasferire AWS ai nostri clienti le efficienze economiche ottenute grazie alla nostra scala. Al 20 settembre 2023, AWS ha ridotto i prezzi 134 volte dal 2006. Se hai ancora qualche dubbio in merito a decisioni commerciali da prendere a causa di questioni relative ai prezzi, puoi fare riferimento ai nuovi tariffari, che includono riduzioni dei prezzi e nuove integrazioni dei servizi. Puoi scoprire le precedenti iniziative di riduzione dei prezzi, comprese le istanze Amazon Elastic Compute Cloud EC2 (Amazon), nella categoria riduzione dei prezzi del News Blog. AWS
- AWS eventi e meetup: partecipa al AWS summit locale e a qualsiasi incontro locale con altre organizzazioni della tua zona. Se non puoi partecipare di persona, prova a partecipare agli eventi virtuali per conoscere meglio gli AWS esperti e i casi aziendali di altri clienti.
- Organizza riunioni con il team del tuo account: pianifica una cadenza regolare di incontri con il team del tuo account, organizza riunioni con il team e discuti delle tendenze del settore e dei servizi AWS . Parla con gli account manager, i solutions architect e i team di supporto a te assegnati.

Risorse

Documenti correlati:

- AWS Gestione dei costi
- Cosa c'è di nuovo con AWS
- AWS Blog di notizie

Esempi correlati:

- AmazonEC2: 15 anni di ottimizzazione e risparmio dei costi IT
- AWS News Blog - Riduzione dei prezzi

COST01-BP08 Creare una cultura attenta ai costi

Implementa modifiche o programmi all'interno dell'organizzazione per creare una cultura consapevole dei costi. Si consiglia di iniziare in piccolo, per poi implementare programmi di grandi dimensioni e di vasta portata all'aumentare delle capacità e dell'utilizzo del cloud da parte dell'organizzazione.

Livello di rischio associato se questa best practice non fosse adottata: basso

Guida all'implementazione

Una cultura consapevole dei costi consente di scalare l'ottimizzazione e la gestione finanziaria del cloud (team operativi e finanziari, centro di eccellenza del cloud, operazioni nel cloud e così via) attraverso best practice eseguite in modo organico e decentralizzato all'interno di tutta l'organizzazione. La consapevolezza dei costi crea livelli elevati di capacità all'interno dell'organizzazione con uno sforzo minimo, qualcosa di analogo a un approccio centralizzato e dall'alto verso il basso.

La creazione della consapevolezza dei costi nel cloud computing, soprattutto per quanto riguarda i principali fattori dei costi, consente ai team di avere la piena consapevolezza dei risultati previsti associati a qualsiasi variazione a livello di costi. I team con accesso agli ambienti cloud devono conoscere i modelli dei prezzi e la differenza tra i tradizionali data center on-premises e il cloud computing.

Il principale vantaggio di una cultura consapevole dei costi è che i team tecnologici ottimizzano i costi in modo proattivo e continuativo (ad esempio, i costi vengono considerati un requisito non funzionale durante la definizione dell'architettura dei nuovi carichi di lavoro oppure quando vengono apportate

modifiche ai carichi di lavoro esistenti) anziché eseguire ottimizzazioni reattive dei costi, in caso di necessità.

Piccoli cambiamenti nella cultura possono avere un grande impatto sull'efficienza dei carichi di lavoro attuali e futuri. Esempi di questo tipo includono:

- Avere visibilità e consapevolezza consente ai team tecnici di progettazione di controllare il loro operato e di capire il tipo di impatto che la loro attività ha in termini di costi.
- Gamificare costi e utilizzo in tutta l'organizzazione. Ciò può essere fatto tramite una dashboard visibile pubblicamente o un report che confronta i costi e l'utilizzo normalizzati tra i team (ad esempio e). cost-per-workload cost-per-transaction
- Premiare l'efficienza dei costi. Ricompensa pubblicamente o privatamente i risultati di ottimizzazione dei costi volontari o non sollecitati e impara dagli errori per evitare di ripeterli in futuro.
- Crea requisiti organizzativi dall'alto verso il basso affinché i carichi di lavoro siano eseguiti nel rispetto dei budget predefiniti.
- Esegui una verifica continua dei requisiti aziendali relativi alle modifiche e dell'impatto dei costi delle modifiche richieste sull'infrastruttura dell'architettura o sulla configurazione del carico di lavoro per avere la certezza di pagare solo quanto è necessario.
- Verifica che il responsabile delle modifiche sia consapevole delle modifiche previste con un impatto sui costi, che a loro volta devono essere confermate dalle parti coinvolte al fine di ottenere risultati aziendali in modo economicamente conveniente.

Passaggi dell'implementazione

- Segnala i costi del cloud ai team tecnologici: per aumentare la consapevolezza dei costi e garantire l'efficienza KPIs per gli stakeholder finanziari e aziendali.
- Comunica le modifiche pianificate alle parti interessate o ai membri dei team: crea una voce nel programma per discutere le modifiche pianificate e l'impatto costi/benefici a livello di carico di lavoro durante le riunioni settimanali.
- Organizza riunioni con il team del tuo account: pianifica una cadenza regolare di incontri con il team del tuo account e discuti delle tendenze del settore e dei servizi AWS . Parla con account manager, architect e team di supporto a te assegnati.
- Condividi storie di successo: condividi storie di successo sulla riduzione dei costi per qualsiasi carico di lavoro o organizzazione per creare un atteggiamento positivo e incoraggiamento verso l'ottimizzazione dei costi. Account AWS

- Formazione: assicurati che i team tecnici o i membri del team siano formati in modo da renderli consapevoli dei costi delle risorse. Cloud AWS
- AWS eventi e incontri: partecipa ai AWS summit locali e a qualsiasi incontro locale con altre organizzazioni della tua zona.
- Iscriviti ai blog: vai alle pagine dei AWS blog e iscriviti al blog What's New e ad altri blog pertinenti per seguire nuove versioni, implementazioni, esempi e modifiche condivise da. AWS

Risorse

Documenti correlati:

- AWS Blog
- AWS Gestione dei costi
- AWS Blog di notizie

Esempi correlati:

- AWS Gestione finanziaria nel cloud
- AWS Well-Architected Labs: gestione finanziaria nel cloud

COST01-BP09 Quantifica il valore aziendale grazie all'ottimizzazione dei costi

La quantificazione del valore aggiunto realizzato tramite l'ottimizzazione dei costi consente di comprendere l'intero set di vantaggi per la tua organizzazione. Poiché l'ottimizzazione dei costi è un investimento necessario, la quantificazione del valore aggiunto consente di spiegare il ritorno sull'investimento alle parti interessate. La quantificazione del valore aggiunto può aiutarti a ottenere maggiori consensi dalle parti interessate sugli investimenti futuri in materia di ottimizzazione dei costi, e fornisce un framework per misurare i risultati delle attività di ottimizzazione dei costi della tua organizzazione.

Livello di rischio associato se questa best practice non fosse adottata: medio

Guida all'implementazione

Quantificare il valore aziendale significa misurare i vantaggi che le aziende ottengono dalle azioni e dalle decisioni che prendono. Il valore aziendale può essere tangibile (riduzione delle spese o

aumento dei profitti) o intangibile (migliore reputazione del marchio o maggiore soddisfazione del cliente).

Quantificare il valore aziendale derivante dall'ottimizzazione dei costi significa determinare il valore o i vantaggi ottenuti dall'impegno dedicato a rendere più efficiente la spesa. Ad esempio, se un'azienda spende 100.000 dollari per implementare un carico di lavoro AWS e successivamente lo ottimizza, il nuovo costo diventa di soli 80.000 dollari senza sacrificare la qualità o l'output. In questo scenario, il valore aziendale quantificato derivante dall'ottimizzazione dei costi è un risparmio di 20.000 dollari. Ma oltre ai semplici risparmi, l'azienda potrebbe anche quantificare il valore in termini di tempi di consegna più rapidi, maggiore soddisfazione dei clienti o altre metriche derivanti dall'impegno nell'ambito dell'ottimizzazione dei costi. Le parti interessate devono prendere decisioni in merito al potenziale valore dell'ottimizzazione dei costi, al costo dell'ottimizzazione del carico di lavoro e al valore del ritorno sugli investimenti.

Oltre a rendicontare i risparmi derivanti dall'ottimizzazione dei costi, è consigliabile quantificare il valore aggiunto fornito. I vantaggi dell'ottimizzazione dei costi sono in genere quantificati in termini di costi inferiori per ottenere un risultato aziendale. Ad esempio, puoi quantificare Amazon Elastic Compute Cloud(AmazonEC2) i risparmi sui costi acquistando Savings Plans, che riducono i costi e mantengono i livelli di output del carico di lavoro. Puoi quantificare le riduzioni dei costi di AWS spesa quando le istanze EC2 Amazon inattive vengono rimosse o i volumi Amazon Elastic Block Store (EBSAmazon) non collegati vengono eliminati.

I vantaggi derivanti dall'ottimizzazione dei costi, tuttavia, vanno oltre la riduzione o l'eliminazione dei costi. Prendi in considerazione l'acquisizione di dati aggiuntivi per misurare i miglioramenti dell'efficienza e il valore aggiunto.

Passaggi dell'implementazione

- Valuta i vantaggi aziendali: questo è il processo di analisi e regolazione dei Cloud AWS costi in modo da massimizzare il beneficio ricevuto da ogni dollaro speso. Invece di concentrarti sulla riduzione dei costi senza considerare il valore aziendale, nell'ambito dell'ottimizzazione dei costi valuta i vantaggi aziendali e il ritorno sugli investimenti, che potrebbero aumentare il valore del denaro speso. Si tratta di spendere con saggezza e di fare investimenti e spese nelle aree che producono i migliori rendimenti.
- Analisi AWS dei costi di previsione: le previsioni aiutano gli stakeholder finanziari a stabilire le aspettative con gli altri stakeholder interni ed esterni dell'organizzazione e possono migliorare la prevedibilità finanziaria dell'organizzazione. AWS Cost Explorerpuò essere utilizzato per eseguire previsioni relative ai costi e all'utilizzo.

Risorse

Documenti correlati:

- Cloud AWS Economia
- AWS Blog
- AWS Gestione dei costi
- AWS Blog di notizie
- Whitepaper sul pilastro dell'affidabilità Well-Architected
- AWS Cost Explorer

Video correlati:

- Sblocca il valore aziendale con Windows on AWS

Esempi correlati:

- Measuring and Maximizing the Business Value of Customer 360
- The Business Value of Adopting Amazon Web Services Managed Databases
- The Business Value of Amazon Web Services for Independent Software Vendors
- Business Value of Cloud Modernization
- The Business Value of Migration to Amazon Web Services

Comprensione delle spese e dell'utilizzo

Questions

- COST2. In che modo gestisci l'utilizzo?
- COST3. In che modo monitori i costi e l'utilizzo?
- COST4. In che modo disattivi le risorse?

COST2. In che modo gestisci l'utilizzo?

Stabilisci policy e meccanismi per verificare che i costi sostenuti mentre raggiungi gli obiettivi siano adeguati. Utilizzando un checks-and-balances approccio, è possibile innovare senza spendere troppo.

Best practice

- COST02-BP01 Sviluppa politiche basate sui requisiti della tua organizzazione
- COST02-BP02 Implementare obiettivi e traguardi
- COST02-BP03 Implementare una struttura dei conti
- COST02-BP04 Implementazione di gruppi e ruoli
- COST02-BP05 Implementare il controllo dei costi
- COST02-BP06 Tieni traccia del ciclo di vita del progetto

COST02-BP01 Sviluppa politiche basate sui requisiti della tua organizzazione

Sviluppa policy che definiscano il modo in cui le risorse vengono gestite dalla tua organizzazione e controllale periodicamente. Le policy devono coprire gli aspetti dei costi relativi alle risorse e ai carichi di lavoro, comprese la creazione, la modifica e la disattivazione nel ciclo di vita delle risorse.

Livello di rischio associato se questa best practice non fosse adottata: elevato

Guida all'implementazione

Comprendere i costi e i fattori chiave della tua organizzazione è fondamentale per gestire i costi e l'utilizzo in modo efficiente e per identificare le opportunità di riduzione dei costi. In genere, le organizzazioni gestiscono molteplici carichi di lavoro eseguiti da più team. Questi team possono trovarsi in diverse unità dell'organizzazione, ciascuna con un proprio flusso di ricavi. La capacità di attribuire i costi delle risorse ai singoli proprietari del carico di lavoro, del prodotto o dell'organizzazione incoraggia un comportamento di utilizzo efficiente e contribuisce a ridurre gli sprechi. Un monitoraggio accurato dei costi e dell'utilizzo consente di comprendere quanto sia ottimizzato un carico di lavoro e quanto siano redditizi i prodotti e le unità organizzative. Questa conoscenza consente di prendere decisioni più informate su dove allocare le risorse all'interno dell'organizzazione. La consapevolezza dell'utilizzo a tutti i livelli dell'organizzazione è fondamentale per promuovere il cambiamento, poiché la modifica dell'utilizzo determina variazioni dei costi. Prova ad adottare una strategia versatile per acquisire consapevolezza delle tue spese.

Il primo passo per attuare la governance consiste nell'utilizzare i requisiti della tua organizzazione per sviluppare policy per l'utilizzo del cloud. Queste policy definiscono il modo in cui l'organizzazione utilizza il cloud e il modo in cui le risorse vengono gestite. Le policy devono coprire tutti gli aspetti dei costi relativi alle risorse e ai carichi di lavoro correlati a costi o utilizzo, compresa la creazione, la modifica e la disattivazione durante il ciclo di vita di una risorsa. Verifica che policy e procedure vengano eseguite e implementate per qualsiasi modifica apportata in un ambiente cloud. Durante gli

incontri per la gestione delle modifiche IT, poni domande relative all'impatto sui costi delle modifiche pianificate (se implicano un aumento o una riduzione), alla giustificazione aziendale e ai risultati attesi.

Le policy devono essere semplici, in modo che siano facilmente comprensibili e possano essere implementate in modo efficace in tutta l'organizzazione. Le policy devono anche essere facili da seguire e interpretare (in modo da essere utilizzate) e specifiche (senza interpretazioni errate tra i team). Inoltre, devono essere ispezionate periodicamente (come i nostri meccanismi) e aggiornate man mano che le condizioni o le priorità aziendali dei clienti cambiano, il che renderebbe la policy obsoleta.

Inizia con policy ampie e di alto livello, ad esempio in quale regione geografica è consentito l'utilizzo o l'ora del giorno in cui le risorse devono essere in esecuzione. Affina gradualmente le policy per le varie unità organizzative e i diversi carichi di lavoro. Le policy comuni includono i servizi e le funzionalità che possono essere utilizzati (ad esempio, archiviazione dalle prestazioni inferiori negli ambienti di test e sviluppo), i tipi di risorse che possono essere utilizzati dai diversi gruppi (ad esempio, le dimensioni massime di una risorsa in un account di sviluppo possono essere impostate su medie) e per quanto tempo queste risorse saranno in uso (temporaneamente, a breve termine o per un periodo di tempo specifico).

Esempio di policy

Di seguito è riportato un esempio di policy che puoi esaminare per creare le tue policy di governance del cloud, basate sull'ottimizzazione dei costi. Assicurati di adattare la policy ai requisiti della tua organizzazione e alle richieste delle parti interessate.

- Nome della policy: definisci un nome chiaro per la policy, ad esempio Ottimizzazione delle risorse e Policy di riduzione dei costi.
- Scopo: spiega perché questa policy dovrebbe essere utilizzata e qual è il risultato previsto. L'obiettivo di questa policy è verificare che sia richiesto un costo minimo per implementare ed eseguire il carico di lavoro desiderato per soddisfare i requisiti aziendali.
- Ambito: definisci chiaramente chi deve utilizzare questa politica e quando deve essere utilizzata, ad esempio DevOps X Team per utilizzarla nei clienti degli Stati Uniti orientali per l'ambiente X (di produzione o non produzione).

Dichiarazione delle policy

1. Seleziona us-east-1 o più regioni Stati Uniti-Est in base all'ambiente del carico di lavoro e ai requisiti aziendali (sviluppo, test di accettazione da parte degli utenti, preproduzione o produzione).
2. Pianifica l'esecuzione delle RDS istanze Amazon EC2 e Amazon tra le sei del mattino e le otto di sera (Eastern Standard Time (EST)).
3. Blocca tutte le EC2 istanze Amazon inutilizzate dopo otto ore e le istanze RDS Amazon non utilizzate dopo 24 ore di inattività.
4. Termina tutte le EC2 istanze Amazon non utilizzate dopo 24 ore di inattività in ambienti non di produzione. Ricorda al proprietario dell'EC2istanza Amazon (in base ai tag) di esaminare le EC2 istanze Amazon interrotte in produzione e informalo che le EC2 sue istanze Amazon verranno chiuse entro 72 ore se non sono in uso.
5. Usa famiglie e dimensioni di istanze generiche, ad esempio m5.large, quindi ridimensiona l'istanza in base all'utilizzo della memoria utilizzata. CPU AWS Compute Optimizer
6. Assegna la priorità utilizzando il dimensionamento automatico per regolare dinamicamente il numero di istanze in esecuzione in base al traffico.
7. Usa le istanze spot per carichi di lavoro non critici.
8. Esamina i requisiti di capacità per impegnare piani di risparmio o istanze riservate per carichi di lavoro prevedibili e informa il team della gestione finanziaria del cloud.
9. Utilizza le policy Amazon S3 del ciclo di vita per spostare i dati a cui si accede di rado su livelli di archiviazione più economici. Se non è stata definita alcuna policy di conservazione, utilizza il Piano intelligente Amazon S3 per spostare automaticamente gli oggetti nel livello archiviato.
10. Monitora l'utilizzo delle risorse e imposta allarmi per attivare eventi di scalabilità utilizzando Amazon. CloudWatch
11. Per ciascuno di essi Account AWS, puoi Budget AWS impostare i budget di costo e utilizzo per il tuo account in base al centro di costo e alle unità aziendali.
12. L'utilizzo Budget AWS per impostare i budget di costo e utilizzo per il tuo account può aiutarti a tenere sotto controllo le spese ed evitare fatture impreviste, consentendoti di controllare meglio i costi.

Procedura: fornisci procedure dettagliate per l'attuazione di questa policy o fai riferimento ad altri documenti che descrivono come implementare ciascuna dichiarazione della policy. Questa sezione dovrebbe fornire step-by-step istruzioni per adempiere ai requisiti della politica.

Per implementare questa politica, è possibile utilizzare vari strumenti o AWS Config regole di terze parti per verificare la conformità alla dichiarazione politica e attivare azioni correttive automatiche

utilizzando le AWS Lambda funzioni. È inoltre possibile utilizzarla AWS Organizations per far rispettare la politica. Inoltre, dovresti controllare regolarmente l'utilizzo delle risorse e modificare la policy, se necessario, per verificare che continui a soddisfare le esigenze aziendali.

Passaggi dell'implementazione

- Incontra le parti interessate: per sviluppare le policy, chiedi alle parti interessate (ufficio aziendale per il cloud, ingegneri o responsabili delle decisioni funzionali per l'applicazione delle policy) all'interno della tua organizzazione di specificare i loro requisiti e documentarli. Segui un approccio iterativo iniziando in modo generale e perfezionando continuamente le unità più piccole in ogni fase. I membri del team includono quelli con interesse diretto nel carico di lavoro, ad esempio unità organizzative o proprietari di applicazioni, nonché gruppi di supporto, come i team di sicurezza e i team finanziari.
- Ottieni conferma: verifica che i team siano d'accordo sulle policy a cui possono accedere e che possono distribuire sull' Cloud AWS. Verifica che rispettino le policy della tua organizzazione e conferma che le creazioni di risorse siano in linea con le policy e le procedure concordate.
- Organizza sessioni di formazione per l'onboarding: chiedi ai nuovi membri dell'organizzazione di partecipare a corsi di formazione di onboarding per sviluppare una consapevolezza sui costi e sui requisiti aziendali. Potrebbero adottare policy diverse legate all'esperienza precedente o non rifletterci affatto.
- Definisci le posizioni del tuo carico di lavoro: definisci dove opera il carico di lavoro, incluso il Paese e l'area all'interno del Paese. Queste informazioni vengono utilizzate per la mappatura Regioni AWS e le zone di disponibilità.
- Definisci e raggruppa servizi e risorse: definisci i servizi necessari per il carico di lavoro. Per ogni servizio, specifica i tipi, la dimensione e il numero di risorse richieste. Definisci i gruppi per le risorse in base alla funzione, ad esempio i server di applicazioni o lo storage di database. Le risorse possono appartenere a più gruppi.
- Definisci e raggruppa gli utenti per funzione: definisci gli utenti che interagiscono con il carico di lavoro, concentrandoti su ciò che fanno e su come utilizzano il carico di lavoro, non su chi sono o sulla loro posizione nell'organizzazione. Raggruppa utenti o funzioni simili. È possibile utilizzare le politiche AWS gestite come guida.
- Definisci le operazioni: utilizzando le posizioni, le risorse e gli utenti identificati in precedenza, definisci le azioni richieste da ciascuno di essi per ottenere i risultati del carico di lavoro durante il ciclo di vita (sviluppo, funzionamento e disattivazione). Identifica le operazioni in base ai gruppi, non ai singoli elementi nei gruppi, in ogni posizione. Inizia in generale con lettura o scrittura, quindi perfeziona le azioni specifiche per ciascun servizio.

- Definisci il periodo di revisione: carichi di lavoro e requisiti organizzativi possono subire modifiche nel tempo. Definisci la pianificazione della revisione del carico di lavoro per assicurarti che sia allineata alle priorità organizzative.
- Documenta le policy: verifica che le policy definite siano accessibili secondo le esigenze dell'organizzazione. Queste policy vengono utilizzate per implementare, mantenere e controllare l'accesso agli ambienti.

Risorse

Documenti correlati:

- Change Management in the Cloud
- AWS Politiche gestite per Job Functions
- AWS strategia di fatturazione per più account
- Azioni, risorse e chiavi di condizione per i servizi AWS
- AWS Gestione e governance
- Controlla l'accesso all' Regioni AWS utilizzo IAM delle politiche
- Infrastrutture globali Regioni e AZs

Video correlati:

- AWS Gestione e governance su larga scala

Esempi correlati:

- VMware- Cosa sono le politiche cloud?

COST02-BP02 Implementare obiettivi e traguardi

Implementa obiettivi e target di costi e utilizzo per il carico di lavoro. Gli obiettivi forniscono indicazioni alla tua organizzazione sui risultati attesi, mentre i target forniscono risultati misurabili per i tuoi carichi di lavoro.

Livello di rischio associato se questa best practice non fosse adottata: elevato

Guida all'implementazione

Sviluppa obiettivi e target di costi e utilizzo per la tua organizzazione. In quanto organizzazione in crescita AWS, è importante stabilire e monitorare gli obiettivi per l'ottimizzazione dei costi. Questi obiettivi o indicatori chiave di prestazione (KPIs) possono includere elementi come la percentuale di spesa su richiesta o l'adozione di determinati servizi ottimizzati come le istanze AWS Graviton o i tipi di volume EBS gp3. La definizione di obiettivi misurabili e raggiungibili ti aiuta a calcolare i miglioramenti dell'efficienza, un fattore importante per le operazioni aziendali. Gli obiettivi forniscono all'organizzazione linee guida e indicazioni sui risultati previsti.

I target forniscono i risultati specifici e misurabili da raggiungere. In breve, un obiettivo è la direzione in cui si vuole andare e un obiettivo indica fino a che punto in quella direzione e quando tale obiettivo dovrebbe essere raggiunto (utilizzare linee guida specifiche, misurabili, assegnabili, realistiche e tempestive, oppure). SMART Un esempio di obiettivo è che l'utilizzo della piattaforma aumenti in modo significativo, con solo un piccolo incremento (non lineare) dei costi. Un esempio di target è un aumento del 20% dell'utilizzo della piattaforma, con un incremento dei costi inferiore al 5%. Un altro obiettivo comune è che i carichi di lavoro devono essere più efficienti ogni sei mesi. L'obiettivo corrispondente prevede che il costo per metrica aziendale debba diminuire del cinque per cento ogni sei mesi. Utilizzate le metriche giuste e impostate i valori calcolati KPIs per la vostra organizzazione. Puoi iniziare con quelli di base KPIs ed evolverli in un secondo momento in base alle esigenze aziendali.

Un obiettivo per l'ottimizzazione dei costi è l'incremento dell'efficienza del carico di lavoro, ossia la riduzione del costo per ogni risultato aziendale del carico di lavoro nel corso del tempo. Implementa questo obiettivo per tutti i carichi di lavoro e stabilisci un target come l'incremento dell'efficienza del 5% ogni 6-12 mesi. Nel cloud, puoi raggiungere questo target attraverso la definizione della capacità di ottimizzazione dei costi, nonché nuove versioni di servizi e funzionalità.

I target sono i benchmark quantificabili che desideri raggiungere per conseguire i tuoi obiettivi e che confrontano i tuoi risultati effettivi rispetto al target. Stabilisci benchmark KPIs per il costo unitario dei servizi di elaborazione (come l'adozione di Spot, l'adozione di Graviton, i tipi di istanze più recenti e la copertura On-Demands), dei servizi di storage (come EBS GP3 adozione, istantanee obsolete EBS e storage standard Amazon S3) o dell'utilizzo dei servizi di database (RDScome motori open source, adozione di Graviton e copertura on demand). Questi benchmark KPIs possono aiutarti a verificare che utilizzi i servizi nel modo più conveniente. AWS

La tabella seguente fornisce un elenco di AWS metriche standard di riferimento. Ogni organizzazione può avere valori target diversi per questi KPIs obiettivi.

Categoria	KPI (%)	Descrizione
Calcolo	EC2Copertura d'uso	EC2istanze (in termini di costi o ore) che utilizzano SP+RI +Spot rispetto al totale (in termini di costi o ore) delle istanze EC2
Calcolo	Utilizzo SP/RI di calcolo	Ore SP o RI utilizzate a fronte delle ore SP o RI totali disponibili
Calcolo	EC2Costo orario/ora	EC2costo diviso per il numero di EC2 istanze in esecuzione in quell'ora
Calcolo	v costo CPU	Costo per v CPU per tutte le istanze
Calcolo	Generazione di istanze più recenti	Percentuale di istanze su Graviton (o altri tipi di istanze di generazione moderna)
Database	RDScopertura	RDSistanze (in termini di costi o ore) che utilizzano RI rispetto al totale (in termini di costi o ore) delle RDS istanze
Database	RDSutilizzo	Ore SP o RI utilizzate rispetto alle ore SP o RI totali disponibi li
Database	RDSoperatività	RDScosto diviso per il numero di RDS istanze in esecuzione in quell'ora

Categoria	KPI (%)	Descrizione
Database	Generazione di istanze più recenti	Percentuale di istanze su Graviton (o altri tipi di istanze moderne)
Storage	Utilizzo dell'archiviazione	Costo dell'archiviazione ottimizzato (ad esempio Glacier, Deep Archive o Infrequent Access) diviso per il costo totale della stessa
Assegnazione di tag	Risorse prive di tag	Esploratore dei costi: 1. Filtra crediti, sconti, tasse, rimborsi, marketplace e copia l'ultimo costo mensile. 2. Seleziona Mostra solo risorse prive di tag in Cost Explorer 3. Dividi l'importo delle risorse prive di tag per il costo mensile.

Utilizzando questa tabella, stabilisci i valori target o benchmark che devono essere calcolati in base agli obiettivi dell'organizzazione. È necessario misurare determinate metriche aziendali e comprendere i risultati aziendali per quel carico di lavoro per definirlo in modo accurato e realistico. KPIs Quando valuti le metriche delle prestazioni di un'organizzazione, tieni in considerazione i vari tipi di metrica che servono a scopi diversi. Queste metriche misurano principalmente le prestazioni e l'efficienza dell'infrastruttura tecnica piuttosto che direttamente l'impatto aziendale complessivo. Ad esempio, possono tenere traccia dei tempi di risposta del server, della latenza della rete o dei tempi di attività del sistema. Queste metriche sono fondamentali per valutare in che misura l'infrastruttura supporta le operazioni tecniche dell'organizzazione. Tuttavia, non forniscono approfondimenti diretti sugli obiettivi aziendali più ampi, come la soddisfazione del cliente, la crescita dei ricavi o la quota

di mercato. Per acquisire un quadro completo delle prestazioni aziendali, integra queste metriche dell'efficienza con le metriche aziendali strategiche direttamente correlate ai risultati aziendali.

Ottieni una visibilità quasi in tempo reale sulle tue opportunità di risparmio KPIs e sulle relative opportunità di risparmio e monitora i tuoi progressi nel tempo. Per iniziare con la definizione e il monitoraggio degli KPI obiettivi, consigliamo la KPI dashboard di Cloud Intelligence Dashboards (CID). Sulla base dei dati del Cost and Usage Report (CUR), la KPI dashboard fornisce una serie di consigli per l'ottimizzazione dei costiKPIs, con la possibilità di impostare obiettivi personalizzati e monitorare i progressi nel tempo.

Se disponi di altre soluzioni per impostare e monitorare KPI gli obiettivi, assicurati che questi metodi siano adottati da tutte le parti interessate alla gestione finanziaria del cloud della tua organizzazione.

Passaggi dell'implementazione

- Definisci i livelli di utilizzo previsti: parti dai livelli di utilizzo. Coinvolgi i responsabili dell'applicazione, i team di marketing e i team aziendali a livello più ampio per capire quali sono i livelli di utilizzo previsti per il carico di lavoro. Considera in che modo potrà cambiare la domanda dei clienti nel corso del tempo e se ci saranno modifiche dovute a incrementi stagionali o campagne di marketing.
- Definisci risorse e costi del carico di lavoro: una volta definiti i livelli di utilizzo, quantifica le modifiche nelle risorse del carico di lavoro necessarie per soddisfarli. Potresti dover aumentare le dimensioni o il numero di risorse per un componente del carico di lavoro, aumentare il trasferimento dei dati o modificare i componenti del carico di lavoro in un servizio diverso a un livello specifico. Specifica i costi per ciascuno di questi punti e prevedine la variazione in caso di modifica dell'utilizzo.
- Definisci gli obiettivi aziendali: prendendo l'output dalle variazioni previste in termini di utilizzo e costi, combinalo con le modifiche previste nella tecnologia o in qualsiasi programma in esecuzione e sviluppa obiettivi per il carico di lavoro. Gli obiettivi devono riguardare l'utilizzo e il costo, nonché la relazione tra i due. Gli obiettivi devono essere semplici, di alto livello e aiutare le persone a capire cosa si aspetta l'azienda in termini di risultati, come avere la certezza che le risorse non utilizzate rimangano al di sotto di determinati livelli di costo. Non è necessario definire gli obiettivi per ogni tipo di risorsa non utilizzato o definire i costi causati dalle perdite per gli obiettivi e i target. Assicurati che siano disponibili programmi a livello di organizzazione (ad esempio lo sviluppo di competenze come la formazione e l'istruzione), se ci sono variazioni previste dei costi senza variazioni di utilizzo.
- Definisci i target: per ciascuno degli obiettivi definiti, specifica un target misurabile. Se l'obiettivo è aumentare l'efficienza nel carico di lavoro, il target quantifica il miglioramento (generalmente

espresso in risultati aziendali per dollaro speso) e il momento in cui sarà efficace. Ad esempio, potresti definire un obiettivo per ridurre al minimo gli sprechi dovuti al provisioning eccessivo. Con questo obiettivo, il target può stabilire che gli sprechi dovuti al provisioning eccessivo di calcolo nel primo livello dei carichi di lavoro di produzione non superino il 10% del costo di calcolo del livello. Inoltre, un secondo target potrebbe stabilire che gli sprechi dovuti al provisioning eccessivo di calcolo nel secondo livello dei carichi di lavoro di produzione non superino il 5% del costo di calcolo del livello.

Risorse

Documenti correlati:

- AWS managed policies for job functions
- Strategia di fatturazione con account multipli di AWS
- Controlla l'accesso all' Regioni AWS utilizzo IAM delle politiche
- S.M.A.R.T. Goals
- Come monitorare l'ottimizzazione dei costi KPIs con la CID KPI Dashboard

Video correlati:

- Well-Architected Labs: obiettivi e target (Livello 100)

Esempi correlati:

- What is a unit metric?
- Selecting a unit metric to support your business
- Unit metrics in practice – lessons learned
- How unit metrics help create alignment between business functions
- Well-Architected Labs: disattivazione delle risorse (obiettivi e target)
- Well-Architected Labs: tipo, dimensione e numero di risorse (obiettivi e target)

COST02-BP03 Implementare una struttura dei conti

Implementa una struttura di account che si adatta alla tua organizzazione. In questo modo sarà possibile ripartire e gestire i costi in tutta l'organizzazione.

Livello di rischio associato se questa best practice non fosse adottata: elevato

Guida all'implementazione

AWS Organizations consente di crearne di più Account AWS che possono aiutarvi a governare centralmente il vostro ambiente man mano che scalate i carichi di lavoro. AWSÈ possibile modellare la gerarchia organizzativa raggruppandola Account AWS in una struttura di unità organizzative (OU) e creandone di più Account AWS in ciascuna unità organizzativa. Per creare una struttura di account, è necessario decidere innanzitutto quale Account AWS sarà l'account di gestione. Successivamente, è possibile creare nuovi account Account AWS o selezionare account esistenti come account membro in base alla struttura degli account progettata seguendo le best practice relative agli account di gestione e alle best practice relative agli account membro.

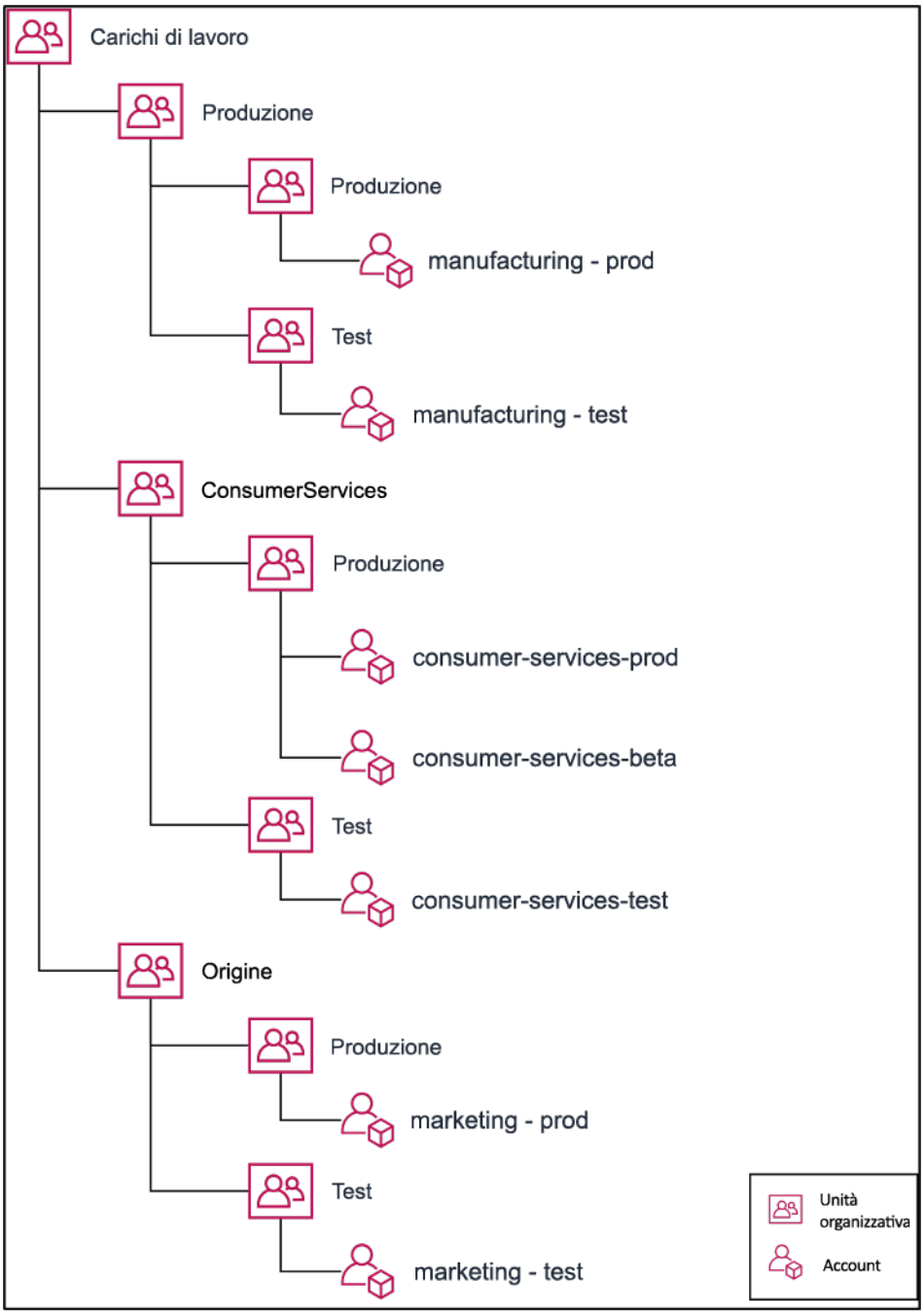
È consigliabile disporre sempre di almeno un account di gestione con un account membro collegato, indipendentemente dalle dimensioni dell'organizzazione o dall'utilizzo. Tutte le risorse del carico di lavoro dovrebbero risiedere solo all'interno degli account membri e nessuna risorsa dovrebbe essere creata all'interno dell'account di gestione. Non esiste una risposta valida per tutti per quanti ne Account AWS dovresti avere. Valuta i tuoi modelli operativi e di costo attuali e futuri per assicurarti che la struttura Account AWS rifletta gli obiettivi della tua organizzazione. Alcune aziende ne creano diversi Account AWS per motivi commerciali, ad esempio:

- È richiesto l'isolamento amministrativo o fiscale e di fatturazione tra unità dell'organizzazione o centri di costo o carichi di lavoro specifici.
- AWS i limiti di servizio sono impostati in modo da essere specifici per determinati carichi di lavoro.
- Esiste un requisito per l'isolamento e la separazione tra carichi di lavoro e risorse.

All'interno di AWS Organizations, la fatturazione consolidata crea il costrutto tra uno o più account membri e l'account di gestione. Gli account membri consentono di isolare e distinguere i costi e l'utilizzo per gruppi. Una pratica comune è quella di avere account membri separati per ciascuna unità dell'organizzazione (come finanza, marketing e vendite), per il ciclo di vita di ciascun ambiente (come sviluppo, test e produzione) o per ciascun carico di lavoro (carico di lavoro a, b e c) e poi aggregare questi account membri tramite la fatturazione consolidata.

La fatturazione consolidata consente di accorpare i pagamenti di più Account AWS membri sotto un unico account di gestione e, al tempo stesso, di fornire comunque visibilità all'attività di ciascun account membro. Il fatto che i costi e l'utilizzo vengono aggregati nell'account di gestione consente di massimizzare gli sconti per volume di servizio e di massimizzare l'utilizzo degli sconti a fronte di impegni (Savings Plans e istanze riservate) per ottenere gli sconti più elevati.

Il diagramma seguente mostra come utilizzare AWS Organizations le unità organizzative (OU) per raggruppare più account e collocarne più di uno Account AWS in ciascuna unità organizzativa. Si consiglia di utilizzarlo OUs per vari casi d'uso e carichi di lavoro e fornisce modelli per l'organizzazione degli account.



Esempio di raggruppamento di più persone Account AWS in unità organizzative.

AWS Control Towerpuò impostare e configurare rapidamente più AWS account, assicurando che la governance sia allineata ai requisiti dell'organizzazione.

Passaggi dell'implementazione

- Definisci i requisiti di separazione: i requisiti di separazione sono una combinazione di più fattori, tra cui sicurezza, affidabilità e costrutti finanziari. Analizza ciascun fattore in ordine e specifica se il carico di lavoro o l'ambiente del carico di lavoro deve essere separato da altri carichi di lavoro. La sicurezza riguarda il rispetto dei requisiti di accesso e di dati. L'affidabilità riguarda la gestione dei limiti, in modo che gli ambienti e i carichi di lavoro non influiscano gli uni sugli altri. Esamina periodicamente i pilastri della sicurezza e dell'affidabilità del Framework Well-Architected e segui le best practice fornite. I costrutti finanziari creano una rigida separazione finanziaria (centri di costo diversi, proprietà e responsabilità dei carichi di lavoro). Esempi comuni di separazione sono i carichi di lavoro di produzione e test eseguiti in account separati o l'utilizzo di un account separato in modo che i dati di fatturazione possano essere forniti ai singoli settori o reparti aziendali dell'organizzazione o alle terze parti proprietarie dell'account.
- Definisci i requisiti di raggruppamento: i requisiti per il raggruppamento non sostituiscono i requisiti di separazione, ma vengono utilizzati a supporto della gestione. Raggruppa ambienti o carichi di lavoro simili che non richiedono separazione. Un esempio è costituito dal raggruppamento di più ambienti di test o sviluppo associati a uno o più carichi di lavoro.
- Definisci la struttura dell'account: utilizzando queste separazioni e questi raggruppamenti, specifica un account per ciascun gruppo e mantieni i requisiti di separazione. Questi account sono i tuoi account membri o collegati. Raggruppando questi account membri in un unico account di gestione o di pagamento, puoi combinarne l'utilizzo, ottenendo maggiori sconti per i volumi e una singola fattura per tutti gli account. È possibile separare i dati di fatturazione e fornire a ciascun account membro una visualizzazione individuale dei dati di fatturazione. Se i dati di utilizzo o di fatturazione di un account membro non devono essere visibili a nessun altro account, o se AWS è richiesta una fattura separata, definisci più account di gestione o di pagamento. In questo caso, ciascun account membro dispone del proprio account di gestione o di pagamento. Le risorse devono sempre essere collocate negli account membri o collegati. Gli account di gestione/di pagamento devono essere utilizzati solo per la gestione.

Risorse

Documenti correlati:

- Utilizzo dei tag per l'allocazione dei costi
- AWS managed policies for job functions
- Strategia di fatturazione con account multipli di AWS

- Controlla l'accesso all'utilizzo delle politiche Regioni AWS IAM
- AWS Control Tower
- AWS Organizations
- Best practice per account di gestione e account membri
- Organizzazione dell' AWS ambiente utilizzando più account
- Attivazione della condivisione di sconti istanze riservate e Savings Plans
- Fatturazione consolidata
- Fatturazione consolidata

Esempi correlati:

- Divisione CUR e condivisione dell'accesso

Video correlati:

- Presentazione AWS Organizations
- Configura un AWS ambiente multi-account che utilizzi le migliori pratiche per AWS Organizations

Esempi correlati:

- Well-Architected Labs: creazione di AWS un'organizzazione (livello 100)
- Divisione e condivisione dell'accesso AWS Cost and Usage Report
- Definizione di una strategia AWS multi-account per le società di telecomunicazioni
- Migliori pratiche per l'ottimizzazione Account AWS
- Migliori pratiche per le unità organizzative con AWS Organizations

COST02-BP04 Implementazione di gruppi e ruoli

Implementa gruppi e ruoli che si allineino alle tue policy e controlla chi può creare, modificare o ritirare istanze e risorse in ogni gruppo. Ad esempio, implementa gruppi di sviluppo, test e produzione. Questo vale per AWS i servizi e le soluzioni di terze parti.

Livello di rischio associato se questa best practice non fosse adottata: basso

Guida all'implementazione

I ruoli e i gruppi di utenti sono elementi costitutivi fondamentali nella progettazione e implementazione di sistemi sicuri ed efficienti. I ruoli e i gruppi aiutano le organizzazioni a trovare il giusto equilibrio a livello di controllo dei requisiti di flessibilità e produttività, supportando in definitiva gli obiettivi organizzativi e le esigenze degli utenti. Come consigliato nella sezione Gestione delle identità e degli accessi di AWS Well-Architected Framework Security Pillar, è necessario disporre di una solida gestione delle identità e delle autorizzazioni per fornire l'accesso alle risorse giuste alle persone giuste nelle giuste condizioni. Gli utenti disporranno solo del livello di accesso necessario per completare le proprie attività. Ciò riduce al minimo il rischio associato all'accesso non autorizzato o all'uso improprio.

Dopo avere sviluppato le policy, è possibile creare gruppi logici e ruoli degli utenti all'interno dell'organizzazione. Ciò consente di assegnare le autorizzazioni, controllare l'utilizzo e semplificare l'implementazione di affidabili meccanismi di controllo degli accessi, impedendo l'accesso non autorizzato alle informazioni sensibili. Inizia con i raggruppamenti di persone di alto livello. Generalmente, questi corrispondono alle unità organizzative e ai ruoli lavorativi (ad esempio, amministratore di sistema nel reparto IT, controllore finanziario o business analyst). I gruppi classificano le persone che eseguono attività simili e necessitano di un accesso simile. I ruoli definiscono che cosa un gruppo deve fare. È più facile gestire le autorizzazioni per gruppi e ruoli che per i singoli utenti. I ruoli e i gruppi assegnano le autorizzazioni in modo coerente e sistematico a tutti gli utenti, evitando errori e incongruenze.

Quando il ruolo di un utente cambia, gli amministratori possono modificare l'accesso a livello di ruolo o di gruppo, anziché riconfigurare i singoli account utente. Ad esempio, un amministratore di sistema nel reparto IT deve disporre di un accesso che permetta di creare tutte le risorse, mentre un membro del team di analisi ha la necessità di creare soltanto risorse di analisi.

Passaggi dell'implementazione

- Implementazione dei gruppi: utilizzando i gruppi di utenti definiti nelle policy dell'organizzazione, implementa i gruppi corrispondenti, se necessario. Per le migliori pratiche su utenti, gruppi e autenticazione, consulta il Security Pillar del AWS Well-Architected Framework.
- Implementazione di ruoli e policy: utilizzando le operazioni definite nelle policy dell'organizzazione, crea le policy di accesso e i ruoli necessari. Per le best practice su ruoli e policy, consulta il Security Pillar del AWS Well-Architected Framework.

Risorse

Documenti correlati:

- AWS managed policies for job functions
- Strategia di fatturazione con account multipli di AWS
- AWS Pilastro di sicurezza Well-Architected Framework
- AWS Identity and Access Management (IAM)
- AWS Identity and Access Management politiche

Video correlati:

- Why use Identity and Access Management

Esempi correlati:

- Well-Architected Labs: Identità e accesso base
- Controlla l'accesso all' Regioni AWS utilizzo IAM delle politiche
- Starting your Cloud Financial Management journey: Cloud cost operations

COST02-BP05 Implementare il controllo dei costi

Implementa controlli basati sulle policy dell'organizzazione e sui gruppi e sui ruoli definiti. Questi garantiscono che i costi siano sostenuti solo in base ai requisiti dell'organizzazione come, ad esempio, il controllo dell'accesso alle regioni o ai tipi di risorse.

Livello di rischio associato se questa best practice non fosse adottata: medio

Guida all'implementazione

Un primo passo comune per implementare i controlli dei costi consiste nell'impostare delle notifiche quando si verificano eventi di costi o utilizzo al di fuori delle policy. In questo modo è possibile agire rapidamente e verificare se è necessaria un'azione correttiva, senza limitare o influire negativamente sui carichi di lavoro o sulle nuove attività. Dopo aver conosciuto i limiti del carico di lavoro e dell'ambiente, puoi applicare la governance. Budget AWSconsente di impostare notifiche e definire budget mensili per AWS i costi, l'utilizzo e gli sconti sugli impegni (Savings Plans e Reserved

Instances). È possibile creare budget a livello di costo aggregato (ad esempio, tutti i costi) o a un livello più granulare, includendo solo dimensioni specifiche come account membri, servizi, tag o zone di disponibilità.

Una volta impostati i limiti di budget Budget AWS, utilizzali AWS Cost Anomaly Detectionper ridurre i costi imprevisti. AWS Cost Anomaly Detection è un servizio di gestione dei costi che utilizza l'apprendimento automatico per monitorare continuamente costi e utilizzo al fine di rilevare spese insolite. Aiuta a individuare le spese anomale e le cause principali in modo da garantire un intervento tempestivo. Innanzitutto, create un sistema di monitoraggio dei costi AWS Cost Anomaly Detection, quindi scegliete la preferenza di notifica impostando una soglia in dollari (ad esempio un avviso in caso di anomalie con impatto superiore a 1.000 USD). Una volta ricevuti gli avvisi, puoi analizzare la causa alla base dell'anomalia e l'impatto sui costi. Puoi inoltre monitorare ed eseguire la tua analisi delle anomalie in AWS Cost Explorer.

Applica le politiche di governance in AWS through AWS Identity and Access Managemente AWS Organizations Service Control Policies (). SCP IAMconsente di gestire in modo sicuro l'accesso a AWS servizi e risorse. UtilizzandoIAM, puoi controllare chi può creare o gestire AWS risorse, il tipo di risorse che possono essere create e dove possono essere create. In questo modo riduci al minimo la possibilità che vengano create risorse al di fuori della policy definita. Utilizza i ruoli e i gruppi creati in precedenza e assegna IAMpolitiche per imporre l'utilizzo corretto. SCPoffre il controllo centralizzato sulle autorizzazioni massime disponibili per tutti gli account dell'organizzazione, garantendo che gli account rispettino le linee guida per il controllo degli accessi. SCPssono disponibili solo in un'organizzazione che ha tutte le funzionalità attivate e puoi configurarle SCPs per negare o consentire azioni per gli account dei membri per impostazione predefinita. Per ulteriori dettagli sull'implementazione della gestione degli accessi, consulta il whitepaper sul pilastro della sicurezza Well-Architected

La governance può essere implementata anche tramite la gestione delle quote di servizio di AWS. Assicurandoti che le quote di servizio siano impostate con spese minime e siano gestite in modo accurato, puoi ridurre al minimo la creazione di risorse che non rientrano nei requisiti della tua organizzazione. Per ottenere questo risultato, devi comprendere la velocità con cui i tuoi requisiti possono cambiare, valutare i progetti in corso (sia la creazione sia la disattivazione di risorse) e considerare la velocità con cui è possibile implementare le modifiche alle quote. Le quote di servizio possono essere utilizzate per aumentare le quote all'occorrenza.

Passaggi dell'implementazione

- Implementa notifiche sulle spese: tramite le policy aziendali definite, crea Budget AWS per ricevere notifiche quando le spese ricadono al di fuori delle tue policy. Configura più budget

dei costi, uno per ogni account, in modo da ricevere informazioni sulla spesa complessiva del conto. Quindi configura budget di costo aggiuntivi all'interno di ciascun account per unità più piccole al suo interno. Queste unità variano a seconda della struttura dell'account. Alcuni esempi comuni sono i Regioni AWS carichi di lavoro (che utilizzano i tag) o AWS i servizi. Configura un elenco di distribuzione e-mail come destinatario per le notifiche e non un account e-mail di una singola persona. È possibile configurare un budget effettivo per quando un importo viene superato oppure utilizzare un budget previsto per la notifica dell'utilizzo previsto. Puoi anche preconfigurare AWS Budget Actions in grado di applicare SCP politiche IAM o politiche specifiche o bloccare istanze Amazon o EC2 Amazon RDS target. Le operazioni di budget possono essere avviate automaticamente o richiedere l'approvazione del flusso di lavoro.

- Implementa notifiche sulle spese anomale: usa AWS Cost Anomaly Detection per ridurre i costi imprevisti dell'organizzazione e analizzare la causa principale delle potenziali spese anomale. Una volta creato Cost Monitor per identificare le spese insolite con la granularità specificata e aver configurato le notifiche AWS Cost Anomaly Detection, ti invia un avviso quando viene rilevata una spesa insolita. Questo ti permetterà di analizzare le cause alla base dell'anomalia e di valutarne l'impatto sui costi. Utilizza AWS Cost Categories durante la configurazione AWS Cost Anomaly Detection per identificare quale team di progetto o team di business unit può analizzare la causa principale del costo imprevisto e intraprendere le azioni necessarie tempestive.
- Implementa i controlli sull'utilizzo: utilizzando le politiche organizzative definite, implementa IAM politiche e ruoli per specificare quali azioni gli utenti possono eseguire e quali azioni non possono. È possibile includere più politiche organizzative in una AWS politica. Nello stesso modo in cui hai definito le policy, inizia in modo generale e quindi applica controlli più dettagliati a ogni fase. Anche le restrizioni dei servizi sono un controllo efficace sull'utilizzo. Implementa le restrizioni dei servizi corrette su tutti gli account.

Risorse

Documenti correlati:

- AWS managed policies for job functions
- Strategia di fatturazione con account multipli di AWS
- Controlla l'accesso all' Regioni AWS utilizzo IAM delle politiche
- Budget AWS
- AWS Cost Anomaly Detection
- Controllate i AWS costi

Video correlati:

- Come posso utilizzarlo Budget AWS per tenere traccia delle mie spese e del mio utilizzo

Esempi correlati:

- Esempi di politiche di gestione degli IAM accessi
- Example service control policies
- AWS Budget e azioni
- Crea una IAM policy per controllare l'accesso alle EC2 risorse Amazon utilizzando i tag
- Limita l'accesso di IAM Identity a EC2 risorse Amazon specifiche
- Crea una IAM politica per limitare l'EC2utilizzo di Amazon per famiglia
- Well-Architected Labs: costi e governance d'uso (Livello 100)
- Well-Architected Labs: costi e governance d'uso (Livello 200)
- Integrazioni Slack per il rilevamento delle anomalie dei costi utilizzando AWS Chatbot

COST02-BP06 Tieni traccia del ciclo di vita del progetto

Rileva, misura e controlla il ciclo di vita di progetti, team e ambienti per evitare di usare risorse non necessarie e pagare per esse.

Livello di rischio associato se questa best practice non fosse adottata: basso

Guida all'implementazione

Monitorando efficacemente il ciclo di vita del progetto, le organizzazioni possono ottimizzare il controllo dei costi attraverso una migliore pianificazione, gestione e ottimizzazione delle risorse. Gli approfondimenti acquisiti attraverso il monitoraggio sono preziosi per la formulazione di decisioni informate che contribuiscono alla competitività a livello di costi e al successo complessivo del progetto.

Il monitoraggio dell'intero ciclo di vita del carico di lavoro consente di capire quando i carichi di lavoro o i suoi componenti non sono più necessari. I carichi di lavoro e i componenti esistenti possono sembrare in uso, ma quando vengono AWS rilasciati nuovi servizi o funzionalità, possono essere smantellati o adottati. Controlla le fasi precedenti dei carichi di lavoro. Quando un carico di lavoro arriva in produzione, gli ambienti precedenti possono essere disattivati o notevolmente ridotti in termini di capacità fino a quando non sono nuovamente necessari.

Puoi applicare i tag alle risorse con un intervallo di tempo o un promemoria per aggiungere l'ora in cui il carico di lavoro è stato esaminato. Ad esempio, se sono trascorsi mesi dall'ultima volta che l'ambiente di sviluppo è stato esaminato, potrebbe essere il momento giusto per esaminarlo nuovamente per verificare se è possibile adottare nuovi servizi o se l'ambiente è in uso. È possibile raggruppare e contrassegnare le applicazioni con myApplicationson AWS per gestire e tenere traccia di metadati quali criticità, ambiente, ultima revisione e centro di costo. Puoi tenere traccia del ciclo di vita del carico di lavoro e monitorare e gestire i costi, lo stato di integrità, il livello di sicurezza e le prestazioni delle applicazioni.

AWS fornisce vari servizi di gestione e governance che è possibile utilizzare per il monitoraggio del ciclo di vita delle entità. È possibile utilizzare il AWS Confignostro AWS Systems Manager per fornire un inventario dettagliato delle AWS risorse e della configurazione. Ti consigliamo di integrare questi servizi con i sistemi di gestione di progetti o asset esistenti per tenere traccia dei progetti attivi e dei prodotti all'interno della tua organizzazione. La combinazione del sistema attuale con il ricco set di eventi e metriche fornito da AWS consente di creare una visione degli eventi significativi del ciclo di vita e di gestire in modo proattivo le risorse per ridurre i costi non necessari.

Analogamente alla gestione del ciclo di vita delle applicazioni (ALM), il monitoraggio del ciclo di vita del progetto dovrebbe comportare la collaborazione di più processi, strumenti e team, ad esempio progettazione e sviluppo, test, produzione, supporto e ridondanza dei carichi di lavoro.

Monitorando attentamente ogni fase del ciclo di vita di un progetto, le organizzazioni ottengono informazioni cruciali e un maggiore controllo, semplificando la pianificazione, l'implementazione e la riuscita del progetto. Questa attenta supervisione verifica che i progetti non solo soddisfino gli standard di qualità, ma vengano consegnati in tempo e nel rispetto del budget, promuovendo l'efficienza complessiva dei costi.

Per ulteriori dettagli sull'implementazione del monitoraggio del ciclo di vita delle entità, consulta il whitepaper sul pilastro dell'eccellenza operativa di AWS Well-Architected.

Passaggi dell'implementazione

- Stabilisci il processo di monitoraggio del ciclo di vita del progetto: il team Centro di eccellenza del cloud deve predisporre un processo di monitoraggio del ciclo di vita del progetto. Stabilisci un approccio strutturato e sistematico per il monitoraggio dei carichi di lavoro al fine di migliorare il controllo, la visibilità e le prestazioni dei progetti. Rendi il processo di monitoraggio trasparente, collaborativo e incentrato sul miglioramento continuo per massimizzarne l'efficacia e il valore.
- Esegui le revisioni del carico di lavoro: in base a quanto definito dalle policy organizzative, stabilisci una cadenza regolare per l'audit dei progetti esistenti e le revisioni del carico di lavoro. L'impegno

speso per il controllo deve essere proporzionale al rischio, al valore o al costo approssimativo per l'organizzazione. Le aree chiave da includere nell'audit sono il rischio di incidente o interruzione per l'organizzazione, il valore o contributo all'organizzazione (misurato in fatturato o reputazione del marchio), il costo del carico di lavoro (misurato come costo totale delle risorse e costi operativi) e l'utilizzo del carico di lavoro (misurato in numero di risultati dell'organizzazione per unità di tempo). Se queste aree cambiano durante il ciclo di vita, sono necessarie modifiche al carico di lavoro, ad esempio la disattivazione completa o parziale.

Risorse

Documenti correlati:

- Linee guida per l'etichettatura AWS
- Che cos'è ALM (Application Lifecycle Management)?
- AWS managed policies for job functions

Esempi correlati:

- Controlla l'accesso all'utilizzo delle politiche Regioni AWS IAM

Strumenti correlati

- AWS Config
- AWS Systems Manager
- Budget AWS
- AWS Organizations
- AWS CloudFormation

COST3. In che modo monitori i costi e l'utilizzo?

Stabilisci policy e procedure per monitorare e allocare i costi in modo appropriato. Ciò ti permette di misurare e migliorare l'efficienza in termini di costi del carico di lavoro.

Best practice

- COST03-BP01 Configurare fonti di informazioni dettagliate
- COST03-BP02 Aggiungere informazioni sull'organizzazione a costi e utilizzo

- COST03-BP03 Identificare le categorie di attribuzione dei costi
- COST03-BP04 Stabilire metriche organizzative
- COST03-BP05 Configurare gli strumenti di fatturazione e gestione dei costi
- COST03-BP06 Alloca i costi in base alle metriche del carico di lavoro

COST03-BP01 Configurare fonti di informazioni dettagliate

Imposta gli strumenti di gestione e report dei costi per ottenere una migliore analisi e trasparenza dei dati sui costi e sull'utilizzo. Configura il carico di lavoro per creare voci di log che facilitino il monitoraggio e la segregazione dei costi e dell'utilizzo.

Livello di rischio associato se questa best practice non fosse adottata: elevato

Guida all'implementazione

Informazioni dettagliate sulla fatturazione, come la granularità oraria negli strumenti di gestione dei costi, consentono alle organizzazioni di tenere traccia dei propri consumi con ulteriori dettagli e di aiutarle a identificare alcuni dei motivi di aumento dei costi. Queste origini dati forniscono la visualizzazione più accurata dei costi e dell'utilizzo dell'intera organizzazione.

È possibile utilizzare Esportazioni di dati AWS per creare esportazioni di () 2.0. AWS Cost and Usage Report CUR Questo è il modo nuovo e consigliato per ricevere i dati dettagliati su costi e utilizzo da AWS. Fornisce informazioni dettagliate sull'utilizzo giornaliero o orario, tariffe, costi e attributi di utilizzo per tutti i AWS servizi a pagamento (le stesse informazioni diCUR), oltre ad alcuni miglioramenti. Tutte le dimensioni possibili includono tag, posizione, attributi delle risorse e account. CUR IDs

Esistono tre tipi di esportazione in base al tipo di esportazione che desideri creare: un'esportazione di dati standard, un'esportazione in una dashboard di costi e utilizzo con QuickSight integrazione Amazon o un'esportazione di dati legacy.

- Esportazione dati standard: esportazione personalizzata di una tabella distribuita su Amazon S3 su base ricorrente.
- Dashboard di costi e utilizzo: esportazione e integrazione in Amazon QuickSight per implementare una dashboard di costi e utilizzo preconfigurata.
- Esportazione dei dati precedenti: esportazione dei dati legacy AWS Cost and Usage Report ()CUR.

È possibile creare esportazioni di dati con le seguenti personalizzazioni:

- Includi risorsa IDs
- Dati di allocazione dei costi suddivisi
- Granularità oraria
- Controllo delle versioni
- Tipo di compressione e formato del file

Per i carichi di lavoro che eseguono container su Amazon ECS o AmazonEKS, abilita i dati di allocazione dei costi suddivisi in modo da poter allocare i costi dei container a singole unità aziendali e team, in base al modo in cui i carichi di lavoro dei container consumano risorse di calcolo e memoria condivise. I dati di allocazione dei costi suddivisi introducono dati su costi e utilizzo per nuove risorse a livello di container. AWS Cost and Usage Report I dati suddivisi sull'allocazione dei costi vengono calcolati calcolando il costo dei singoli ECS servizi e attività in esecuzione sul cluster.

Una dashboard di costi e utilizzo esporta la tabella del dashboard di costi e utilizzo in un bucket S3 su base ricorrente e distribuisce una dashboard di costi e utilizzo preconfigurata su Amazon. QuickSight Utilizza questa opzione se desideri implementare rapidamente una dashboard dei dati su costi e utilizzo senza funzionalità di personalizzazione.

Se lo desideri, puoi comunque esportare CUR in modalità legacy, dove puoi integrare altri servizi di elaborazione, AWS Gluead esempio preparare i dati per l'analisi ed eseguire analisi dei dati con Amazon Athena utilizzando SQL per interrogare i dati.

Passaggi dell'implementazione

- Crea esportazioni di dati: crea esportazioni personalizzate con i dati che desideri e controlla lo schema delle tue esportazioni. Crea esportazioni di dati di fatturazione e gestione dei costi utilizzando Basic SQL e visualizza i tuoi dati di fatturazione e gestione dei costi integrandoti con Amazon. QuickSight Puoi anche esportare i dati in modalità standard per analizzarli con altri strumenti di elaborazione, come Amazon Athena.
- Configura il report su costi e utilizzo: utilizzando la console di fatturazione, configura almeno un report costi e utilizzo. Configura un report con granularità oraria che includa tutti gli identificatori e le risorse. IDs Puoi anche creare altri report con granularità diverse per fornire informazioni di riepilogo di livello superiore.
- Configura la granularità oraria in Cost Explorer: per accedere ai dati su costi e utilizzo con granularità oraria negli ultimi 14 giorni, prendi in considerazione l'abilitazione di dati a livello di ora e risorsa nella console di fatturazione.

- Configura la registrazione dei log da parte delle applicazioni: verifica che l'applicazione registri ogni risultato aziendale che distribuisce in modo che possa essere monitorato e misurato. Assicurati che la granularità di questi dati sia almeno oraria, affinché possa essere abbinata ai dati relativi a costi e utilizzo. Per maggiori dettagli su creazione di log e monitoraggio, consulta il pilastro dell'eccellenza operativa Well-Architected.

Risorse

Documenti correlati:

- Esportazioni di dati AWS
- AWS Glue
- Amazon QuickSight
- AWS Cost Management Pricing
- Applicazione di tag alle risorse AWS
- Analyzing your costs with Cost Explorer
- Managing AWS Cost and Usage Report s
- Pilastro dell'eccellenza operativa Well-Architected

Esempi correlati:

- Configurazione dell'account AWS
- Esportazioni di dati per la AWS fatturazione e la gestione dei costi
- AWS Cost Explorer Casi di utilizzo comune

COST03-BP02 Aggiungere informazioni sull'organizzazione a costi e utilizzo

Definisci uno schema per l'applicazione di tag in base alla tua organizzazione, agli attributi del carico di lavoro e alle categorie di allocazione dei costi, in modo da poter filtrare e cercare le risorse o monitorare costi e utilizzo negli strumenti di gestione dei costi. Implementa un'applicazione di tag consistente in tutte le risorse possibili per scopo, team, ambiente o altri criteri pertinenti alla tua azienda.

Livello di rischio associato se questa best practice non fosse adottata: medio

Guida all'implementazione

Implementa l'applicazione di tag AWS in per aggiungere informazioni sull'organizzazione alle risorse, che verranno, quindi, integrate alle informazioni su costi e utilizzo. Un tag è una coppia chiave-valore: la chiave è definita e deve essere univoca all'interno dell'organizzazione, mentre il valore è univoco per un gruppo di risorse. Ad esempio, una coppia chiave-valore può essere costituita da `Environment`, con un valore di `Production`. Tutte le risorse nell'ambiente di produzione avranno questa coppia chiave-valore. L'applicazione di tag consente di categorizzare e monitorare i costi con informazioni significative e rilevanti sull'organizzazione. Puoi applicare tag che rappresentano categorie dell'organizzazione (ad esempio, centri di costo, nomi di applicazioni, progetti o proprietari) e identificano carichi di lavoro e rispettive funzionalità (come test o produzione) per attribuire i costi e l'utilizzo all'interno dell'organizzazione.

Quando applichi tag alle tue AWS risorse (come Amazon Elastic Compute Cloud istanze o Amazon Simple Storage Service bucket) e attivi i tag, AWS aggiunge queste informazioni ai report sui costi e sull'utilizzo. Puoi creare report e condurre analisi su risorse con tag e senza tag per incrementare la conformità con le policy di gestione dei costi interne e garantire un'attribuzione accurata.

La creazione e l'implementazione di uno standard di AWS etichettatura negli account dell'organizzazione consente di gestire e governare AWS gli ambienti in modo coerente e uniforme. Utilizza le politiche sui tag AWS Organizations per definire le regole su come i tag possono essere utilizzati sulle AWS risorse dei tuoi account in. AWS Organizations Le politiche sui tag consentono di adottare facilmente un approccio standardizzato per AWS etichettare le risorse

AWS Tag Editor consente di aggiungere, eliminare e gestire i tag di più risorse. Con questa funzionalità, è possibile cercare le risorse a cui applicare tag e quindi gestirli per quelle risorse dei tuoi risultati di ricerca.

AWS Cost Categories consente di assegnare un significato organizzativo ai costi, senza richiedere tag sulle risorse. Puoi mappare le informazioni su costi e utilizzo attribuendole a strutture organizzative interne univoche. Puoi definire regole di categoria per mappare e categorizzare i costi utilizzando le dimensioni di fatturazione, ad esempio account e tag. Questo offre un altro livello di funzionalità di gestione oltre all'applicazione di tag. Puoi anche mappare account e tag specifici attribuendoli a più progetti.

Passaggi dell'implementazione

- Definisci uno schema per l'applicazione di tag: riunisci tutte le parti interessate dell'azienda per definire uno schema. Questo generalmente include i ruoli tecnici, finanziari e di gestione.

Definisci un elenco di tag che tutte le risorse devono avere, nonché un elenco di tag che le risorse dovrebbero avere. Verifica che i nomi e i valori dei tag siano coerenti all'interno dell'organizzazione.

- Risorse di tag: utilizzando le categorie di attribuzione dei costi definite, posiziona i tag in tutte le risorse dei carichi di lavoro in base alle categorie. Utilizza strumenti come Tag Editor o AWS Systems Manager per aumentare l'efficienza. CLI
- AWS Implementazione delle categorie di costo: è possibile creare categorie di costi senza implementare l'etichettatura. Le categorie di costo utilizzano le dimensioni di costo e utilizzo esistenti. Crea regole di categoria dallo schema e implementale nelle categorie di costo.
- Automatizza l'applicazione dei tag: per verificare di mantenere elevati livelli di applicazione di tag tra tutte le risorse, automatizza l'applicazione di tag in modo che le risorse siano contrassegnate automaticamente al momento della creazione. Utilizza servizi come AWS CloudFormation per verificare l'avvenuta applicazione di tag alle risorse al momento della creazione. Puoi anche creare una soluzione personalizzata per l'applicazione in automatico di tag mediante le funzioni Lambda o usare un microservizio che scansioni periodicamente il carico di lavoro e rimuova le risorse prive di tag, l'ideale per ambienti di test e sviluppo.
- Monitora ed elabora report sull'applicazione di tag: per verificare di mantenere elevati livelli di applicazione di tag nella tua organizzazione, elabora report e monitora i tag tra i tuoi carichi di lavoro. Puoi utilizzare AWS Cost Explorer per visualizzare il costo delle risorse con tag e senza tag oppure utilizzare servizi come Tag Editor. Verifica regolarmente il numero di risorse senza tag e aggiungi i tag fino a raggiungere il livello desiderato di applicazione di tag.

Risorse

Documenti correlati:

- Tagging Best Practices
- AWS CloudFormation Tag di risorsa
- Cost Categories AWS
- Risorse di etichettatura AWS
- Analisi dei costi con Budgets AWS
- Analyzing your costs with Cost Explorer
- Gestione del report su costi e utilizzo di AWS

Video correlati:

- Come posso etichettare AWS le mie risorse per suddividere la fattura per centro di costo o progetto
- Risorse per l'etichettatura AWS

COST03-BP03 Identificare le categorie di attribuzione dei costi

Identifica le categorie dell'organizzazione, come business unit, reparti o progetti, che potrebbero essere utilizzate per allocare i costi alle entità responsabili dei consumi interni. Utilizza queste categorie per rafforzare la responsabilità della spesa, creare consapevolezza dei costi e promuovere comportamenti di consumo efficaci.

Livello di rischio associato se questa best practice non fosse adottata: elevato

Guida all'implementazione

Il processo di categorizzazione dei costi è fondamentale nella definizione del budget, nella contabilità, nella rendicontazione finanziaria, nel processo decisionale, nell'analisi comparativa e nella gestione dei progetti. La categorizzazione e la classificazione delle spese consentono ai team di comprendere meglio i tipi di costi che dovranno sostenere durante il loro percorso verso il cloud, aiutandoli a prendere decisioni informate e a gestire i budget in modo efficace.

La responsabilità della spesa cloud costituisce un forte incentivo per una gestione disciplinata della domanda e dei costi. Il risultato è un notevole risparmio sui costi del cloud per le organizzazioni che destinano la maggior parte della loro spesa per il cloud a business unit o team che utilizzano il cloud. Inoltre, l'allocazione della spesa per il cloud aiuta le organizzazioni ad adottare un numero maggiore di best practice di governance del cloud centralizzate.

Collabora con il tuo team finanziario e altre parti interessate per comprendere i requisiti di allocazione dei costi all'interno della tua organizzazione durante le chiamate organizzate con periodicità regolare. I costi del carico di lavoro devono essere allocati per tutto il ciclo di vita, inclusi sviluppo, test, produzione e disattivazione. Comprendi in che modo i costi sostenuti per formazione, sviluppo del personale e creazione di idee sono attribuiti all'interno dell'organizzazione. Questo può essere utile per assegnare correttamente gli account utilizzati a questo scopo ai budget di formazione e sviluppo anziché ai budget di costi IT generici.

Dopo aver definito le categorie di attribuzione dei costi con le parti interessate dell'organizzazione, utilizza AWS Cost Categories per raggruppare le informazioni su costi e utilizzo in categorie significative Cloud AWS, ad esempio il costo di un progetto specifico o Account AWS per reparti o unità aziendali. Puoi creare categorie personalizzate e mappare le informazioni su costi e utilizzo

in queste categorie in base a regole che definisci usando componenti diversi come account, tag, servizio o tipo di addebito. Una volta impostate le categorie di costi, è possibile visualizzare le informazioni su costi e utilizzo consentendo così alla tua organizzazione di prendere decisioni di acquisto e strategiche migliori. Queste categorie sono visibili anche in AWS Cost Explorer Budget AWS, e AWS Cost and Usage Report .

Ad esempio, create categorie di costi per le vostre unità aziendali (DevOps team) e per ogni categoria create più regole (regole per ogni sottocategoria) con più dimensioni (tag di allocazione dei costi Account AWS, servizi o tipo di addebito) in base ai raggruppamenti definiti. Con le categorie di costo puoi organizzare i costi utilizzando un motore basato su regole. Le regole configurate organizzeranno i costi in categorie. All'interno di queste regole, puoi filtrare utilizzando più dimensioni per ogni categoria, ad esempio specifiche Account AWS, AWS servizi o tipi di addebito. È possibile utilizzare queste categorie tra più prodotti nelle console di AWS Billing and Cost Management e Cost Management. Ciò include AWS Cost Explorer Budget AWS, AWS Cost and Usage Report, e AWS Cost Anomaly Detection.

Come esempio, il diagramma seguente mostra in che modo raggruppare i costi e le informazioni sull'utilizzo nella tua organizzazione definendo più team (categoria di costo), molteplici ambienti (regole) e assegnando a ogni ambiente molteplici risorse o asset (dimensioni).

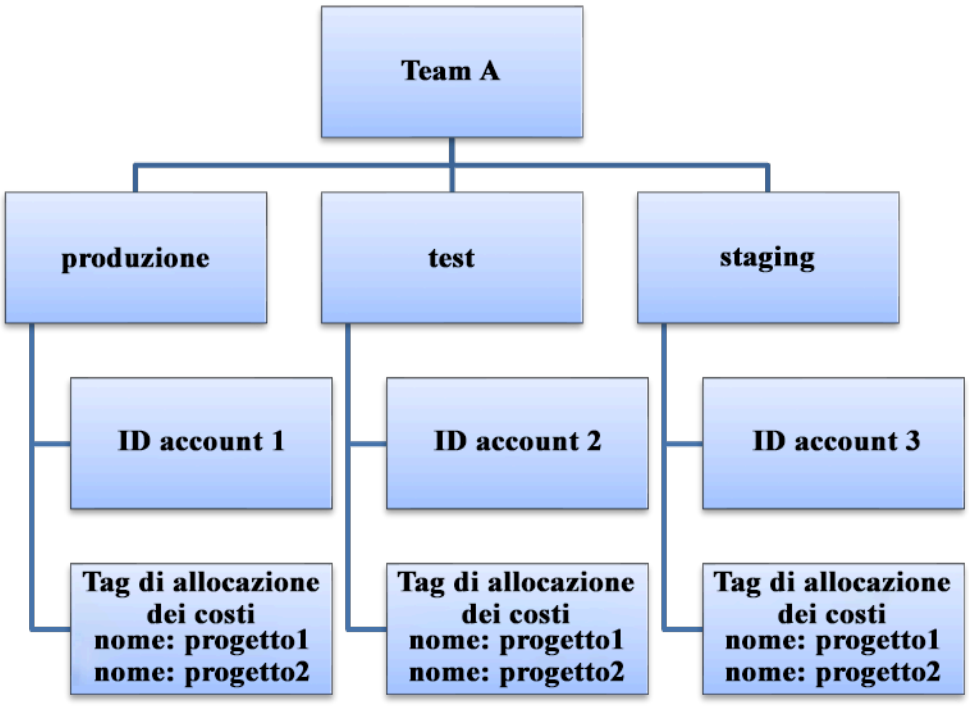


Diagramma relativo a costi e utilizzo

Puoi anche creare gruppi di costi con le categorie di costo. Dopo aver creato le categorie di costo (attendi fino a 24 ore dopo la creazione di una categoria di costo affinché i record di utilizzo siano aggiornati con i valori), verranno visualizzati in AWS Cost Explorer, Budget AWS, AWS Cost and Usage Report e AWS Cost Anomaly Detection. In AWS Cost Explorer and Budget AWS, una categoria di costo viene visualizzata come dimensione di fatturazione aggiuntiva. Puoi utilizzare questa opzione per filtrare il valore specifico della categoria di costo o raggruppare in base alla categoria di costo.

Passaggi dell'implementazione

- Definisci le categorie dell'organizzazione: organizza riunioni con le parti interessate interne e le business unit per definire categorie che riflettano la struttura e i requisiti della tua organizzazione. Queste categorie dovrebbero corrispondere direttamente alla struttura delle categorie finanziarie esistenti, ad esempio business unit, budget, centro di costi o reparto. Osserva i risultati che il cloud offre per la tua azienda, ad esempio la formazione o l'istruzione, poiché anche queste sono categorie organizzative.
- Definisci le categorie funzionali: organizza riunioni con le parti interessate interne e le unità di business per definire categorie che riflettano le funzioni presenti all'interno della tua azienda. Potrebbe trattarsi del carico di lavoro o dei nomi delle applicazioni e del tipo di ambiente, ad esempio produzione, test o sviluppo.
- Definizione AWS delle categorie di costo: crea categorie di costi per organizzare le informazioni su AWS costi e utilizzo utilizzando Cost Categories e mappare AWS costi e utilizzo in categorie significative. A una risorsa possono essere assegnate più categorie e una risorsa può essere in più categorie diverse, quindi definisci tutte le categorie necessarie in modo da essere in grado di gestire i tuoi costi all'interno delle strutture categorizzate mediante le categorie di costo AWS .

Risorse

Documenti correlati:

- Applicazione di tag alle risorse AWS
- Utilizzo dei tag per l'allocazione dei costi
- Analisi dei costi con Budget AWS
- Analyzing your costs with Cost Explorer
- Managing AWS Cost and Usage Report s
- Cost Categories AWS

- Gestione dei costi con AWS Cost Categories
- Creating cost categories
- Tagging cost categories
- Splitting charges within cost categories
- Funzionalità delle categorie di costo AWS

Esempi correlati:

- Organizza i dati relativi a costi e utilizzo con AWS Cost Categories
- Gestione dei costi con AWS Cost Categories
- Well-Architected Labs: visualizzazione di costi e utilizzo
- Well-Architected Labs: categorie di costo

COST03-BP04 Stabilire metriche organizzative

Definisci i parametri dell'organizzazione necessari per il carico di lavoro. I parametri esemplificativi di un carico di lavoro sono i report dei clienti prodotti o le pagine Web scaricate dai clienti.

Livello di rischio associato se questa best practice non fosse adottata: elevato

Guida all'implementazione

Comprendi in che modo viene misurato l'output del carico di lavoro rispetto al successo aziendale. Ogni carico di lavoro ha in genere un piccolo set di output principali che indicano le prestazioni. Se disponi di un carico di lavoro complesso con molti componenti, puoi dare priorità alle voci dell'elenco o definire e monitorare i parametri per ogni componente. Collabora con i tuoi team per capire quali parametri utilizzare. Questa unità verrà utilizzata per comprendere l'efficienza del carico di lavoro o il costo per ciascun output aziendale.

Passaggi dell'implementazione

- Definisci i risultati del carico di lavoro: organizza riunioni con le parti interessate dell'azienda e definisci i risultati del carico di lavoro. Si tratta di una misura principale dell'utilizzo da parte dei clienti e devono essere parametri aziendali e non parametri tecnici. Deve esserci un piccolo numero di parametri di alto livello (meno di cinque) per carico di lavoro. Se il carico di lavoro produce più risultati per diversi casi d'uso, raggruppali in un singolo parametro.

- Definisci i risultati dei componenti del carico di lavoro: facoltativamente, se disponi di un carico di lavoro grande e complesso oppure puoi suddividere facilmente il carico di lavoro in componenti (ad esempio microservizi) con input e output ben definiti, definisci i parametri per ogni componente. Lo sforzo deve riflettere il valore e il costo del componente. Inizia con i componenti più grandi e punta ai componenti più piccoli.

Risorse

Documenti correlati:

- Etichettare AWS le risorse
- Analisi dei costi con Budgets AWS
- Analyzing your costs with Cost Explorer
- Gestione del report su costi e utilizzo di AWS

COST03-BP05 Configurare gli strumenti di fatturazione e gestione dei costi

Configura gli strumenti di gestione dei costi in conformità alle policy della tua organizzazione per gestire e ottimizzare gli investimenti nel cloud. Sono inclusi servizi, strumenti e risorse per organizzare e monitorare i dati su costi e utilizzo, migliorare il controllo tramite la fatturazione consolidata e le autorizzazioni di accesso, perfezionare la pianificazione tramite budget e previsioni, ricevere notifiche o avvisi e ridurre i costi tramite l'ottimizzazione di prezzi e risorse.

Livello di rischio associato se questa best practice non fosse adottata: elevato

Guida all'implementazione

Per definire consapevolezza e responsabilità forti, la strategia che interessa l'account deve essere considerata come parte integrante della strategia di allocazione dei costi. Definisci questo concetto ora per non doverlo affrontare in futuro. In caso contrario, il livello di consapevolezza potrebbe essere insufficiente e potrebbero verificarsi problemi in seguito.

Per incoraggiare la responsabilità degli investimenti nel cloud, fornisci agli utenti l'accesso a strumenti che offrono visibilità su costi e utilizzo. AWS consiglia di configurare tutti i carichi di lavoro e definire i team per i seguenti scopi:

- Organizzazione: definisci l'allocazione dei costi e i riferimenti della governance con la tua strategia di applicazione dei tag e la tassonomia. Crea più AWS account con strumenti come o Organization.

AWS Control Tower AWS Etichetta le AWS risorse supportate e classificale in modo significativo in base alla struttura dell'organizzazione (unità aziendali, reparti o progetti). Etichetta i nomi degli account per centri di costo specifici e mappali con AWS Cost Categories per raggruppare gli account delle unità aziendali nei rispettivi centri di costo in modo che il proprietario dell'unità aziendale possa vedere il consumo di più account in un unico posto.

- Accesso: tieni traccia delle informazioni di fatturazione a livello di organizzazione nella fatturazione consolidata e verifica che le parti interessate e i responsabili idonei abbiano accesso.
- Controllo: crea meccanismi di governance efficaci con le giuste protezioni per prevenire scenari imprevisti quando utilizzi Service Control Policies (SCP), tag policy, IAM policy e avvisi di budget. Ad esempio, puoi consentire ai team di creare risorse specifiche nelle regioni preferite solo utilizzando meccanismi di controllo efficaci e impedire la creazione di risorse prive di tag specifici, come il centro di costo.
- Stato attuale: configura un pannello di controllo che mostra i livelli correnti di costi e utilizzo. Il pannello di controllo deve essere disponibile in un luogo altamente visibile all'interno dell'ambiente di lavoro, in modo simile al pannello di controllo delle operazioni. Puoi esportare i dati e utilizzare la Dashboard costi e utilizzo dalla Centrale ottimizzazione costi AWS o qualsiasi prodotto supportato per creare questa visibilità. Potresti dover creare pannelli di controllo diversi per tipi di utenti diversi, ad esempio il pannello di controllo per i manager sarà diverso da quello di progettazione.
- Notifiche: invia notifiche quando il costo o l'utilizzo superano i limiti definiti e si verificano anomalie con AWS Budgets o Cost Anomaly Detection. AWS
- Report: riepiloga tutte le informazioni su costi e utilizzo. Aumenta la consapevolezza e la responsabilità dei tuoi investimenti nel cloud con dati sui costi dettagliati e attribuibili. Crea i report con i suggerimenti pertinenti per il team che li utilizza.
- Monitoraggio: mostra i costi e l'utilizzo attuali rispetto a obiettivi o target stabiliti.
- Analisi: offri ai membri del team la possibilità di eseguire analisi personalizzate e approfondite fino alla granularità oraria, giornaliera o mensile con diversi filtri (risorse, account, tag, ecc.).
- Esame: non perdere gli aggiornamenti sulle opportunità di implementazione delle risorse e di ottimizzazione dei costi. Ricevi notifiche utilizzando Amazon CloudWatchSNS, Amazon o Amazon SES per la distribuzione delle risorse a livello di organizzazione. Consulta i consigli per l'ottimizzazione dei costi con AWS Trusted Advisor o. AWS Compute Optimizer
- Report delle tendenze: mostra la variabilità dei costi e dell'utilizzo nel periodo richiesto e con la granularità richiesta.
- Previsioni: mostra i costi futuri stimati, prevedi l'utilizzo delle risorse e investi con pannelli di controllo di previsioni che tu stesso crei.

Sfrutta la Centrale ottimizzazione costi AWS per esaminare da una posizione centralizzata le potenziali opportunità di risparmio sui costi consolidati, nonché per creare esportazioni di dati per l'integrazione con Amazon Athena. Puoi anche utilizzare il AWS Cost Optimization Hub per implementare il Cost and Usage Dashboard, che utilizza Amazon QuickSight per l'analisi interattiva dei costi e la condivisione sicura delle informazioni sui costi.

Se non disponi delle competenze o della larghezza di banda essenziali nella tua organizzazione, puoi lavorare con AWS ProServ, AWS Managed Services (AMS) o partner.AWS Puoi anche utilizzare strumenti di terze parti, ma assicurati di convalidare la proposta di valore.

Passaggi dell'implementazione

- Consenti l'accesso agli strumenti in base ai team: configura i tuoi account e crea gruppi con accesso ai report su costi e utilizzo necessari per i loro consumi e usa AWS Identity and Access Management per controllare l'accesso a strumenti come AWS Cost Explorer. Questi gruppi devono includere i rappresentanti di tutti i team che possiedono o gestiscono un'applicazione. In questo modo si certifica che ogni team ha accesso alle informazioni sui costi e sull'utilizzo per tenere traccia dei propri consumi.
- Organizza tag e categorie di costo: organizza i costi tra team, business unit, applicazioni, ambienti e progetti. Usa i tag delle risorse per organizzare i costi, in base ai tag di allocazione dei costi. Crea le categorie di costo in base alle dimensioni utilizzando tag, account, servizi e così via per mappare i costi.
- Configura AWS i budget: configura AWS i budget su tutti gli account per i tuoi carichi di lavoro. Imposta un budget per la spesa complessiva dell'account e un budget per il carico di lavoro utilizzando i tag e le categorie di costo. Configura le notifiche in AWS Budget per ricevere avvisi quando superi gli importi previsti o quando i costi stimati superano i budget.
- Configura AWS il rilevamento delle anomalie nei AWS costi: utilizza Cost Anomaly Detection per i tuoi account, i servizi principali o le categorie di costi che hai creato per monitorare costi e utilizzo e rilevare spese insolite. Puoi ricevere avvisi singolarmente in report aggregati e ricevere avvisi in un'e-mail o in un SNS argomento Amazon, il che ti consente di analizzare e determinare la causa principale dell'anomalia e identificare il fattore che determina l'aumento dei costi.
- Usa gli strumenti di analisi dei costi: configura AWS Cost Explorer per il tuo carico di lavoro e gli account e visualizza i dati sui costi per ulteriori analisi. Crea un pannello di controllo per il carico di lavoro che tenga traccia della spesa generale e le metriche di utilizzo chiave per il carico di lavoro, nonché preveda i costi futuri sulla base dei tuoi dati storici.
- Utilizza strumenti di analisi per il risparmio: utilizza AWS Cost Optimization Hub per identificare le opportunità di risparmio con consigli personalizzati, tra cui l'eliminazione delle risorse inutilizzate,

il dimensionamento corretto, i piani di risparmio, le prenotazioni e i consigli sugli ottimizzatori di calcolo.

- Configura strumenti avanzati: puoi creare in modo facoltativo immagini per agevolare l'analisi interattiva e la condivisione delle informazioni sui costi. Con Data Exports on AWS Cost Optimization Hub, puoi creare un dashboard di costi e utilizzo basato su Amazon QuickSight per la tua organizzazione che fornisce dettagli e granularità aggiuntivi. Puoi anche implementare funzionalità di analisi avanzate utilizzando le esportazioni di dati in Amazon Athena per query avanzate e creare dashboard su Amazon. QuickSight Collabora con i partner AWS per adottare soluzioni di gestione del cloud per il monitoraggio e l'ottimizzazione della fatturazione consolidata del cloud.

Risorse

Documenti correlati:

- Cos'è la gestione dei AWS Billing and Cost Management costi?
- Creazione di un AWS ambiente basato sulle migliori pratiche
- Migliori pratiche per l'etichettatura delle risorse AWS
- Etichettare le tue risorse AWS
- Cost Categories AWS
- Analisi dei costi con Budgets AWS
- Analisi dei costi con AWS Cost Explorer
- Che cos'è l'esportazione AWS dei dati?

Video correlati:

- Deploying Cloud Intelligence Dashboards
- Ricevi avvisi su qualsiasi metrica di ottimizzazione dei costi FinOps o KPI

Esempi correlati:

- Dashboard di costi e utilizzo fornito da Amazon QuickSight
- Workshop su AWS Cost and Usage Governance

COST03-BP06 Alloca i costi in base alle metriche del carico di lavoro

Alloca i costi del carico di lavoro in base alle metriche di utilizzo o ai risultati aziendali per misurare l'efficienza dei costi del carico di lavoro. Implementa un processo per analizzare i dati relativi a costi e utilizzo con i servizi di analisi, che possono fornire informazioni approfondite e funzionalità di chargeback.

Livello di rischio associato se questa best practice non fosse adottata: basso

Guida all'implementazione

Ottimizzare i costi significa conseguire i risultati aziendali al prezzo più basso eseguendo l'allocazione dei costi del carico di lavoro in base alle metriche di quest'ultimo, misurate in termini di efficienza. Monitora le metriche del carico di lavoro definite tramite file di log o altre funzionalità di monitoraggio dell'applicazione. Combina questi dati con i costi del carico di lavoro, che possono essere ottenuti osservando i costi con un determinato valore di tag o ID account. Esegui questa analisi a livello orario. L'efficienza cambia in genere se disponi di componenti di costo statico, come un database backend sempre in esecuzione, con un tasso di richiesta variabile, ad esempio picchi di utilizzo tra le 9:00 e le 17:00 con poche richieste di notte. Comprendere la relazione tra i costi statici e i costi variabili ti aiuterà a rendere più mirate le tue attività di ottimizzazione.

La creazione di parametri del carico di lavoro per risorse condivise può essere difficile rispetto a risorse come le applicazioni containerizzate su Amazon Elastic Container Service (Amazon) ECS e Amazon Gateway. API Tuttavia, esistono alcuni modi per classificare l'utilizzo e tenere traccia dei costi. Se devi tenere traccia di Amazon ECS e delle risorse AWS Batch condivise, puoi abilitare la suddivisione dei dati di allocazione dei costi. AWS Cost Explorer Con i dati di allocazione dei costi suddivisi, puoi analizzare e ottimizzare i costi e l'utilizzo delle tue applicazioni containerizzate e riallocare i costi delle applicazioni alle singole entità aziendali in base al modo in cui vengono consumate le risorse di calcolo e memoria condivise.

Passaggi dell'implementazione

- Alloca i costi alle metriche del carico di lavoro: utilizzando le metriche e l'applicazione di tag definiti e configurati, crea una metrica che combini l'output e il costo del carico di lavoro. Utilizza servizi di analisi come Amazon Athena e Amazon QuickSight per creare una dashboard di efficienza per il carico di lavoro complessivo e tutti i componenti.

Risorse

Documenti correlati:

- Taggare le risorse AWS
- Analisi dei costi con Budgets AWS
- Analyzing your costs with Cost Explorer
- Gestione del report su costi e utilizzo di AWS

Esempi correlati:

- Migliora la visibilità dei costi di Amazon ECS e AWS Batch con AWS Split Cost Allocation Data

COST4. In che modo disattivi le risorse?

Implementa il controllo delle modifiche e la gestione delle risorse dall'inizio del progetto fino a. end-of-life In questo modo, puoi disattivare o terminare le risorse non utilizzate per ridurre gli sprechi.

Best practice

- COST04-BP01 Tieni traccia delle risorse per tutto il loro ciclo di vita
- COST04-BP02 Implementare un processo di smantellamento
- COST04-BP03 Risorse di smantellamento
- COST04-BP04 Disattiva automaticamente le risorse
- COST04-BP05 Applica le politiche di conservazione dei dati

COST04-BP01 Tieni traccia delle risorse per tutto il loro ciclo di vita

Definisci e implementa un metodo per monitorare le risorse e le loro associazioni con i sistemi durante il loro ciclo di vita. Puoi usare l'applicazione di tag per identificare il carico di lavoro o la funzione della risorsa.

Livello di rischio associato se questa best practice non fosse adottata: elevato

Guida all'implementazione

Disattiva le risorse dei carichi di lavoro che non sono più necessarie. Un esempio comune sono le risorse utilizzate per i test: dopo il completamento dei test, le risorse possono essere rimosse. La tracciabilità delle risorse con i tag (e la predisposizione di report su tali tag) può aiutare a identificare le risorse da disattivare, poiché non saranno più in uso o la loro licenza è in scadenza. L'utilizzo dei

tag è un modo efficace per monitorare le risorse: puoi etichettare la risorsa con la relativa funzione o con una data nota in cui può essere disattivata. Puoi quindi eseguire i report su questi tag. Esempi di valori per l'applicazione di tag relativi alle funzionalità sono `feature-X testing` per identificare lo scopo della risorsa in termini di ciclo di vita del carico di lavoro. Un altro esempio è l'utilizzo di `LifeSpan` o `TTL` per le risorse, ad esempio il nome della chiave e il valore del to-be-deleted tag per definire il periodo di tempo o l'ora specifica per la disattivazione.

Passaggi dell'implementazione

- Implementa uno schema di applicazione di tag: implementa uno schema di applicazione di tag che identifichi il carico di lavoro a cui appartiene la risorsa, verificando che tutte le risorse all'interno del carico di lavoro siano contrassegnate di conseguenza. L'applicazione dei tag aiuta a classificare le risorse in base allo scopo, al team, all'ambiente o ad altri criteri rilevanti per l'azienda. Per ulteriori informazioni su casi d'uso, strategie e tecniche di applicazione dei tag, consulta AWS Tagging Best Practices.
- Implementa il monitoraggio di throughput del carico di lavoro o output: implementa il monitoraggio degli allarmi del throughput del carico di lavoro, avviandolo per le richieste di input o i completamenti dell'output. Configuralo per fornire notifiche quando le richieste o gli output del carico di lavoro scendono a zero, indicando che le risorse del carico di lavoro non sono più utilizzate. Incorpora un fattore temporale se il carico di lavoro scende periodicamente a zero in condizioni normali. Per maggiori informazioni sulle risorse inutilizzate o sottoutilizzate, consulta Controlli dell'ottimizzazione dei costi AWS Trusted Advisor.
- AWS Risorse di gruppo: crea gruppi per le AWS risorse. È possibile utilizzare AWS Resource Groupsper organizzare e gestire le AWS risorse che si trovano all'interno delle stesse Regione AWS. Puoi aggiungere tag alla maggior parte delle risorse affinché sia possibile identificarle e ordinarle all'interno dell'organizzazione. Usa l'editor di tag per aggiungere tag alle risorse supportate in blocco. Prendi in considerazione l'utilizzo di AWS Service Catalog per creare, gestire e distribuire agli utenti finali portafogli di prodotti approvati e gestire il ciclo di vita del prodotto.

Risorse

Documenti correlati:

- AWS Auto Scaling
- AWS Trusted Advisor
- AWS Trusted Advisor Controlli di ottimizzazione dei costi
- Etichettatura delle risorse AWS

- Publishing Custom Metrics

Video correlati:

- Come ottimizzare i costi utilizzando AWS Trusted Advisor

Esempi correlati:

- Organizza AWS le risorse
- Ottimizza i costi utilizzando AWS Trusted Advisor

COST04-BP02 Implementare un processo di smantellamento

Implementa un processo per identificare e disattivare le risorse inutilizzate.

Livello di rischio associato se questa best practice non fosse adottata: elevato

Guida all'implementazione

Implementa un processo standardizzato in tutta l'organizzazione per identificare e rimuovere le risorse inutilizzate. Il processo deve definire la frequenza di esecuzione della ricerca e i processi per rimuovere la risorsa al fine di verificare che tutti i requisiti dell'organizzazione siano soddisfatti.

Passaggi dell'implementazione

- Crea e implementa un processo di disattivazione: collaborando con sviluppatori e proprietari del carico di lavoro, crea un processo di disattivazione per il carico di lavoro e le relative risorse. Il processo deve includere il metodo per verificare se il carico di lavoro è in uso e quello per capire se ciascuna delle risorse del carico di lavoro è in uso. Specifica le fasi necessarie per disattivare la risorsa, rimuovendola dal servizio e garantendo allo stesso tempo la conformità a qualsiasi requisito normativo. Dovrebbero essere incluse tutte le risorse associate, come le licenze o lo spazio di archiviazione collegato. Invia una notifica ai proprietari del carico di lavoro indicando che il processo di disattivazione è stato avviato.

  Utilizza i seguenti passaggi di disattivazione per guidarti su quali dovrebbero essere le verifiche eseguite come parte del processo:

  - Identifica le risorse da disattivare: individua le risorse idonee alla disattivazione nel tuo ambiente Cloud AWS. Registra tutte le informazioni necessarie e pianifica la disattivazione. Nella

sequenza temporale, assicurati di tenere conto di eventuali problemi imprevisti e di quando si verificano durante il processo.

- Coordina e comunica: collabora con i proprietari dei carichi di lavoro per ricevere conferma circa le risorse da eliminare.
- Registra metadati e crea backup: registra i metadati (come publicIPs, Region, AZVPC, Subnet e Security Groups) e crea backup (come snapshot o acquisizione di Amazon Elastic Block StoreAMI, esportazione di chiavi ed esportazione di certificati) se sono necessari per le risorse nell'ambiente di produzione o se sono risorse critiche.
- Convalida infrastructure-as-code: determina se le risorse sono state distribuite con AWS CloudFormation Terraform o qualsiasi altro strumento di infrastructure-as-code distribuzione in modo che possano essere ridistribuite se necessario. AWS Cloud Development Kit (AWS CDK)
- Impedisci l'accesso: applica controlli restrittivi per un periodo di tempo, in modo da impedire l'uso delle risorse mentre stabilisci se sono necessarie o meno. Verifica che l'ambiente delle risorse possa essere ripristinato allo stato originale, se necessario.
- Segui il processo di disattivazione interno: segui le attività amministrative e il processo di disattivazione della tua organizzazione, come rimuovere la risorsa dal dominio dell'organizzazione, rimuovere il DNS record e rimuovere la risorsa dallo strumento di gestione della configurazione, dallo strumento di monitoraggio, dallo strumento di automazione e dagli strumenti di sicurezza.

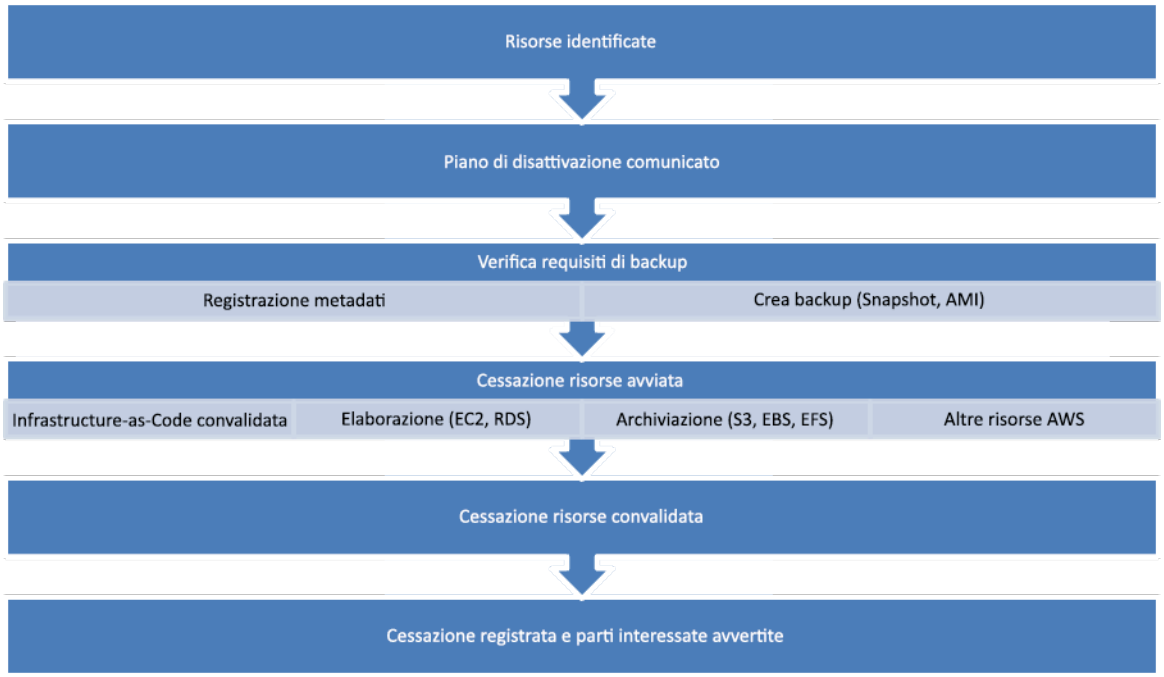
Se la risorsa è un'EC2istanza Amazon, consulta il seguente elenco. Per maggiori dettagli, consulta Come posso eliminare o terminare le mie EC2 risorse Amazon?

- Arresta o interrompi tutte le EC2 istanze e i sistemi di bilanciamento del carico di Amazon. Le EC2 istanze Amazon sono visibili nella console per un breve periodo dopo la loro chiusura. Non verrà addebitato alcun costo per le istanze che non si trovano in stato di esecuzione
- Elimina la tua infrastruttura Auto Scaling.
- Rilascia tutti gli host dedicati.
- Elimina tutti i EBS volumi Amazon e gli EBS snapshot Amazon.
- Rilascia tutti gli Indirizzi IP elastici.
- Annulla la registrazione di tutte le Amazon Machine Images ()AMIs.
- Termina tutti gli ambienti. AWS Elastic Beanstalk

Se la risorsa è un oggetto in uno spazio di archiviazione Amazon S3 Glacier e se si elimina un archivio prima di aver raggiunto la durata minima di archiviazione, verrà addebitato un costo di eliminazione anticipata proporzionale. La durata minima di archiviazione Amazon S3

Glacier dipende dalla classe di archiviazione utilizzata. Per un riepilogo della durata minima dell'archiviazione per ogni classe di archiviazione, consulta Prestazioni delle classi di archiviazione Amazon S3. Per informazioni sulle modalità di calcolo delle tariffe di eliminazione anticipata, consulta i prezzi di Amazon S3.

Il seguente semplice diagramma di flusso del processo di disattivazione illustra le fasi della disattivazione. Prima di disattivare le risorse, verifica che le risorse identificate per la disattivazione non siano utilizzate dall'organizzazione.



Flusso di disattivazione delle risorse.

Risorse

Documenti correlati:

- AWS Auto Scaling
- AWS Trusted Advisor
- AWS CloudTrail

Video correlati:

- Elimina CloudFormation lo stack ma conserva alcune risorse
- Scopri quale utente ha lanciato un'EC2istanza Amazon

Esempi correlati:

- Eliminare o terminare le risorse Amazon EC2
- Scopri quale utente ha lanciato un'EC2istanza Amazon

COST04-BP03 Risorse di smantellamento

Disattiva le risorse attivate da eventi come audit periodici o modifiche relative all'utilizzo. La disattivazione viene in genere eseguita periodicamente e può essere manuale o automatizzata.

Livello di rischio associato se questa best practice non fosse adottata: medio

Guida all'implementazione

La frequenza e lo sforzo di ricerca delle risorse inutilizzate dovrebbero riflettere i risparmi potenziali, pertanto un account con costi contenuti deve essere analizzato con una frequenza minore rispetto a un account che ha costi maggiori. Gli eventi di ricerca e disattivazione possono essere avviati da modifiche di stato nel carico di lavoro, ad esempio il termine del ciclo di vita di un prodotto o la sua sostituzione. Le ricerche e gli eventi di disattivazione possono anche essere avviati da eventi esterni, ad esempio cambiamenti nelle condizioni di mercato o cessazione del prodotto.

Passaggi dell'implementazione

- Disattivazione delle risorse: si tratta della fase di ammortamento delle risorse AWS non più necessarie o al termine di un contratto di licenza. Completa tutti i controlli finali prima di passare alla fase di dismissione e disattivazione delle risorse per evitare interruzioni indesiderate durante fasi come l'esecuzione di snapshot o backup. Utilizzando il processo di disattivazione, disattiva tutte le risorse identificate come inutilizzate.

Risorse

Documenti correlati:

- AWS Auto Scaling
- AWS Trusted Advisor

Esempi correlati:

- Well-Architected Labs: disattivazione delle risorse (Livello 100)

COST04-BP04 Disattiva automaticamente le risorse

Progetta il tuo carico di lavoro in modo da gestire in modo controllato la terminazione delle risorse, identificando e disattivando le risorse non critiche, le risorse non necessarie o quelle a basso utilizzo.

Livello di rischio associato se questa best practice non fosse adottata: basso

Guida all'implementazione

Utilizza l'automazione per ridurre o rimuovere i costi associati al processo di ritiro. Progettare il carico di lavoro per eseguire automaticamente la disattivazione ridurrà i costi complessivi del carico di lavoro durante il suo ciclo di vita. Puoi utilizzare Amazon EC2 Auto Scaling o Application Auto Scaling per eseguire il processo di smantellamento. Puoi anche implementare codice personalizzato utilizzando APIo disattivare automaticamente le risorse del carico SDK di lavoro.

Le applicazioni moderne vengono create innanzitutto in modalità serverless, una strategia che dà priorità all'adozione di servizi serverless. AWS servizi serverless sviluppati per tutti e tre i livelli dello stack: elaborazione, integrazione e archivi dati. L'utilizzo di un'architettura serverless consente di risparmiare sui costi nei periodi di scarso traffico e di approfittare del dimensionamento automatico.

Passaggi dell'implementazione

- EC2Implementa Amazon Auto Scaling o Application Auto Scaling: per le risorse supportate, configurale con Amazon Auto Scaling o Application Auto EC2 Scaling. Questi servizi possono aiutarti a ottimizzare l'utilizzo e l'efficienza dei costi durante l'utilizzo dei servizi. AWS Quando la domanda diminuisce, questi servizi rimuovono automaticamente la capacità di risorse in eccesso per evitare spese inutili.
- Configurazione CloudWatch per terminare le istanze: le istanze possono essere configurate per terminare utilizzando allarmi. CloudWatch Utilizzando i parametri del processo di disattivazione, implementa un allarme con un'operazione Amazon Elastic Compute Cloud. Verifica l'operazione in un ambiente non di produzione prima di eseguire il roll out.
- Implementa il codice all'interno del carico di lavoro: puoi utilizzare o per disattivare le risorse del AWS SDK carico di lavoro. AWS CLI Implementa all'interno dell'applicazione il codice che si integra AWS e termina o rimuove le risorse non più utilizzate.
- Utilizza servizi serverless: dai la priorità alla creazione di architetture serverless e architetture basate sugli eventi per creare ed eseguire le tue applicazioni. AWS AWS offre diversi servizi tecnologici serverless che forniscono intrinsecamente un utilizzo delle risorse ottimizzato automaticamente e uno smantellamento automatizzato (scalabilità in e out). Con le applicazioni

serverless, l'utilizzo delle risorse viene ottimizzato automaticamente e non si paga mai il provisioning in eccesso.

Risorse

Documenti correlati:

- Amazon EC2 Auto Scaling
- Guida introduttiva ad Amazon EC2 Auto Scaling
- Application Auto Scaling
- AWS Trusted Advisor
- Serverless attivo AWS
- Create Alarms to Stop, Terminate, Reboot, or Recover an Instance
- Aggiungere azioni di interruzione agli allarmi Amazon CloudWatch

Esempi correlati:

- Pianificazione dell'eliminazione automatica degli stack AWS CloudFormation
- Well-Architected Labs: disattivazione automatica delle risorse (Livello 100)
- Pulizia automatica serba AWS

COST04-BP05 Applica le politiche di conservazione dei dati

Definisci le policy di conservazione dei dati sulle risorse supportate per gestire l'eliminazione degli oggetti in base ai requisiti della tua organizzazione. Identifica ed elimina risorse non necessarie oppure orfane e oggetti non più richiesti.

Livello di rischio associato se questa best practice non fosse adottata: medio

Usa le policy di conservazione dei dati e del ciclo di vita per ridurre i costi associati al processo di disattivazione e i costi di archiviazione per le risorse identificate. La definizione delle policy di conservazione dei dati e del ciclo di vita per eseguire l'eliminazione e la migrazione di classi di archiviazione automatizzate contribuirà a ridurre i costi di archiviazione generale durante la sua durata. Puoi utilizzare Amazon Data Lifecycle Manager per automatizzare la creazione e l'eliminazione di snapshot di Amazon Elastic Block Store e Amazon Machine Images () con supporto di Amazon, AMIs e utilizzare EBS Amazon S3 Intelligent-Tiering o una configurazione del ciclo di

vita di Amazon S3 per gestire il ciclo di vita dei tuoi oggetti Amazon S3. Puoi anche implementare codice personalizzato utilizzando o per creare policy e regole relative al ciclo di vita per gli oggetti da eliminare automaticamente. API SDK

Passaggi dell'implementazione

- Usa Amazon Data Lifecycle Manager: utilizza le policy del ciclo di vita su Amazon Data Lifecycle Manager per automatizzare l'eliminazione degli snapshot di Amazon con supporto da Amazon. EBS EBS AMIs
- Imposta la configurazione del ciclo di vita su un bucket: usa la configurazione del ciclo di vita di Amazon S3 su un bucket per definire le operazioni che Amazon S3 deve intraprendere durante il ciclo di vita dell'oggetto, oltre all'eliminazione alla fine del ciclo di vita dello stesso, in base ai requisiti aziendali.

Risorse

Documenti correlati:

- AWS Trusted Advisor
- Amazon Data Lifecycle Manager
- Come creare una configurazione del ciclo di vita dei bucket Amazon S3

Video correlati:

- Automatizza Amazon EBS Snapshot con Amazon Data Lifecycle Manager
- Empty an Amazon S3 bucket using a lifecycle configuration rule

Esempi correlati:

- Empty an Amazon S3 bucket using a lifecycle configuration rule
- Well-Architected Lab: disattivazione automatica delle risorse (Livello 100)

Risorse convenienti in termini di costo

Questions

- COST5. In che modo valuti i costi quando selezioni i servizi?

- COST6. In che modo raggiungi gli obiettivi di costo quando selezioni il tipo, le dimensioni e il numero delle risorse?
- COST7. In che modo impieghi i modelli di prezzo per ridurre i costi?
- COST8. In che modo pianifichi i costi per il trasferimento dei dati?

COST5. In che modo valuti i costi quando selezioni i servizi?

Amazon EC2EBS, Amazon e Amazon S3 sono servizi integrati AWS . I servizi gestiti, come Amazon RDS e Amazon DynamoDB, sono servizi di livello superiore o a livello di applicazione. AWS Selezionando i blocchi predefiniti e i servizi gestiti appropriati, è possibile ottimizzare questo carico di lavoro per i costi. Ad esempio, utilizzando i servizi gestiti, puoi ridurre o eliminare gran parte dei costi generali amministrativi e operativi, liberandotene per lavorare su applicazioni e attività correlate al tuo business.

Best practice

- COST05-BP01 Identificare i requisiti organizzativi in termini di costi
- COST05-BP02 Analizza tutti i componenti del carico di lavoro
- COST05-BP03 Eseguire un'analisi approfondita di ogni componente
- COST05-BP04 Selezione del software con licenze convenienti
- COST05-BP05 Seleziona i componenti di questo carico di lavoro per ottimizzare i costi in linea con le priorità dell'organizzazione
- COST05-BP06 Eseguire l'analisi dei costi per usi diversi nel tempo

COST05-BP01 Identificare i requisiti organizzativi in termini di costi

Lavora con i membri del team per definire il bilanciamento tra l'ottimizzazione dei costi e altri pilastri, come le prestazioni e l'affidabilità, per questo carico di lavoro.

Livello di rischio associato se questa best practice non fosse adottata: elevato

Guida all'implementazione

Nella maggior parte delle organizzazioni, il reparto di tecnologia dell'informazione (IT) è composto da diversi team di piccole dimensioni, ciascuno con una propria agenda e area di interesse, che riflettono le specializzazioni e le competenze dei suoi membri. È necessario comprendere gli obiettivi, le priorità e le finalità generali dell'organizzazione e in che modo ogni reparto o progetto contribuisce a

tali obiettivi. La catalogazione di tutte le risorse essenziali, inclusi personale, attrezzature, tecnologia, materiali e servizi esterni, è fondamentale per il raggiungimento degli obiettivi organizzativi e una pianificazione precisa del budget. L'adozione di questo approccio sistematico all'identificazione e alla comprensione dei costi è fondamentale per stabilire un piano dei costi realistico e affidabile per l'organizzazione.

Al momento di selezionare i servizi per un carico di lavoro, è fondamentale comprendere le priorità dell'organizzazione. Crea un equilibrio tra l'ottimizzazione dei costi e altri pilastri del AWS Well-Architected Framework, come prestazioni e affidabilità. È necessario eseguire questo processo sistematicamente e regolarmente in modo da acquisire i cambiamenti a livello di obiettivi, condizioni di mercato e dinamiche operative dell'organizzazione. Un carico di lavoro completamente ottimizzato per i costi è la soluzione più in linea con i requisiti della tua organizzazione, e non necessariamente quella con il costo più basso. Per raccogliere il maggior numero di informazioni, interpella tutti i team all'interno dell'organizzazione, come i team dedicati ai prodotti, di business, tecnici e finanziari. Valuta l'impatto dei compromessi tra interessi concorrenti o approcci alternativi, per aiutare a prendere decisioni informate quando si stabilisce dove concentrare le attività o scegliere una linea di azione.

Ad esempio, accelerare l'introduzione sul mercato di nuove funzionalità può essere preferibile all'ottimizzazione dei costi. Oppure, è possibile scegliere un database relazionale per i dati non relazionali per semplificare la migrazione di un sistema, anziché migrare a un database ottimizzato per il tuo tipo di dati e aggiornare l'applicazione.

Passaggi dell'implementazione

- Identifica i requisiti dell'organizzazione sui costi: organizza riunioni con i membri dei team della tua organizzazione, tra cui i team di gestione dei prodotti, i team proprietari delle applicazioni, i team operativi e di sviluppo, i team di gestione e finanziari. Dai la priorità ai pilastri Well-Architected per questo carico di lavoro e i relativi componenti. L'output dovrebbe essere un elenco ordinato dei pilastri. Puoi anche aggiungere un fattore di ponderazione a ciascun pilastro per indicare il livello di attenzione aggiuntiva assegnato o quanto è simile il livello di attenzione assegnato a due pilastri.
- Analizza il debito tecnico e documentalo: durante la revisione del carico di lavoro, analizza il debito tecnico. Documenta gli elementi lasciati in sospeso per riesaminare il carico di lavoro in un secondo momento, con l'obiettivo di rifattorizzarlo o riprogettarlo per ottimizzarlo ulteriormente. Alle altre parti interessate è fondamentale comunicare in modo chiaro le scelte di compromesso adottate.

Risorse

Best practice correlate:

- REL11-BP07 Progetta il tuo prodotto per soddisfare gli obiettivi di disponibilità e gli accordi sui livelli di servizio di uptime () SLAs
- OPS01-BP06 Valuta i compromessi

Documenti correlati:

- AWS Calcolatore del costo totale di proprietà () TCO
- Classi di archiviazione Amazon S3
- Prodotti cloud

COST05-BP02 Analizza tutti i componenti del carico di lavoro

Verifica che ogni componente del carico di lavoro venga analizzato, indipendentemente dalle dimensioni attuali o dai costi correnti. L'attività di revisione deve riflettere i potenziali benefici, come i costi correnti e quelli previsti.

Livello di rischio associato se questa best practice non fosse adottata: elevato

Guida all'implementazione

I componenti del carico di lavoro, progettati per fornire valore aziendale all'organizzazione, possono includere vari servizi. Per ogni componente, è possibile scegliere Cloud AWS servizi specifici per soddisfare le esigenze aziendali. Questa selezione potrebbe essere influenzata da fattori quali la familiarità o l'esperienza precedente nell'uso di questi servizi.

Dopo aver identificato i requisiti dell'organizzazione come indicato in COST05-BP01 Identifica i requisiti dell'organizzazione in termini di costi, esegui un'analisi approfondita di tutti i componenti del carico di lavoro. Analizza ogni componente considerando i costi e le dimensioni attuali e previsti. Considera il costo dell'analisi rispetto a qualsiasi potenziale risparmio del carico di lavoro durante il suo ciclo di vita. L'impegno dedicato all'analisi di tutti i componenti di questo carico di lavoro deve corrispondere al potenziale risparmio o ai miglioramenti previsti derivanti dall'ottimizzazione del componente specifico. Ad esempio, se il costo della risorsa proposta è di 10 USD al mese e secondo le previsioni i carichi non dovrebbero superare i 15 USD al mese, spendere un giorno di lavoro per ridurre i costi del 50% (5 USD al mese) potrebbe eccedere il potenziale beneficio nel corso della vita del sistema. Usa una stima basata sui dati, più rapida ed efficiente, per generare il migliore risultato complessivo per questo componente.

I carichi di lavoro possono cambiare nel corso del tempo e il giusto set di servizi potrebbe non essere ottimale se l'architettura o l'utilizzo del carico di lavoro cambiano. L'analisi per la selezione dei servizi deve integrare gli stati del carico di lavoro e i livelli di utilizzo attuali e futuri. Implementare un servizio in funzione dello stato o dell'utilizzo futuro del carico di lavoro può ridurre i costi complessivi, diminuendo o rimuovendo l'impegno necessario per apportare modifiche future. Ad esempio, inizialmente l'utilizzo di EMR Serverless potrebbe essere la scelta appropriata. Tuttavia, con l'aumento del consumo per quel servizio, la transizione a EMR on EC2 potrebbe ridurre i costi per quella componente del carico di lavoro.

AWS Cost Explorere AWS Cost and Usage Report s (CUR) può analizzare il costo di un proof of concept (PoC) o di un ambiente di esecuzione. Puoi anche utilizzare AWS Pricing Calculator per stimare i costi del carico di lavoro.

Scrivi un flusso di lavoro che dovrà essere usato dai team tecnici per esaminare i carichi di lavoro. Il flusso di lavoro deve essere semplice, ma coprire tutti i passaggi necessari affinché i team comprendano ogni componente del carico di lavoro e i relativi prezzi. L'organizzazione può quindi seguire e personalizzare il flusso di lavoro in base alle esigenze specifiche di ogni team.

1. Elenca ogni servizio in uso per il tuo carico di lavoro: questa pratica è un buon punto di partenza. Identifica tutti i servizi attualmente in uso e da dove derivano i costi.
2. Scopri come funzionano i prezzi per questi servizi: analizza il modello di prezzo di ciascun servizio. AWS Servizi diversi hanno modelli di prezzo diversi in base a fattori come il volume di utilizzo, il trasferimento dei dati e i prezzi specifici delle funzionalità.
3. Concentrati sui servizi che comportano costi di carico di lavoro imprevisti e che non sono in linea con l'utilizzo previsto e il risultato aziendale: identifica i valori anomali o i servizi il cui costo non è proporzionale al valore o all'utilizzo di or s. AWS Cost Explorer AWS Cost and Usage ReportÈ importante correlare i costi ai risultati aziendali per poter definire le priorità delle attività di ottimizzazione.
4. AWS Cost Explorer, CloudWatch Logs, VPC Flow Logs e Amazon S3 Storage Lens per comprendere la causa principale di questi costi elevati: questi strumenti sono fondamentali nella diagnosi dei costi elevati. Ogni servizio offre una visione diversa per osservare e analizzare l'utilizzo e i costi. Ad esempio, Cost Explorer aiuta a determinare le tendenze complessive dei costi, CloudWatch Logs fornisce informazioni operative, VPC Flow Logs visualizza il traffico IP e Amazon S3 Storage Lens è utile per l'analisi dello storage.
5. Utilizzalo Budget AWS per impostare budget per determinati importi per servizi o account: l'impostazione dei budget è un modo proattivo per gestire i costi. Consente Budget AWS di impostare soglie di budget personalizzate e ricevere avvisi quando i costi superano tali soglie.

6. Configura gli CloudWatch allarmi Amazon per inviare avvisi di fatturazione e utilizzo: configura il monitoraggio e gli avvisi per i parametri di costo e utilizzo. CloudWatch gli allarmi possono avvisarti quando vengono superate determinate soglie, il che migliora i tempi di risposta dell'intervento.

Favorisci notevoli miglioramenti e risparmi finanziari nel tempo con la revisione strategica di tutti i componenti del carico di lavoro, indipendentemente dalle caratteristiche attuali. L'impegno profuso in questo processo di revisione deve essere ponderato, con un'attenta considerazione dei potenziali vantaggi che si possono ottenere.

Passaggi dell'implementazione

- Elenca i componenti del carico di lavoro: crea un elenco dei componenti del carico di lavoro. Usa questo elenco per verificare che ogni componente sia stato analizzato. Gli impegni sostenuti devono riflettere la criticità del carico di lavoro secondo quanto definito dalle priorità dell'organizzazione. Raggruppa le risorse in modo funzionale, ad esempio in base all'archiviazione del database di produzione, per migliorare l'efficienza se sono presenti più database.
- Assegna priorità all'elenco dei componenti: assegna ai componenti nell'elenco una priorità in termini di impegno richiesto. Tale assegnazione viene in genere eseguita in ordine dal componente più costoso a quello meno costoso o in base alla criticità definita dalle priorità dell'organizzazione.
- Esegui l'analisi: per ciascun componente dell'elenco, esamina le opzioni e i servizi disponibili e scegli l'opzione che si allinea meglio alle priorità dell'organizzazione.

Risorse

Documenti correlati:

- AWS Pricing Calculator
- AWS Cost Explorer
- Classi di archiviazione Amazon S3
- Prodotti Cloud AWS

Video correlati:

- AWS Serie di ottimizzazione dei costi: CloudWatch

COST05-BP03 Eseguire un'analisi approfondita di ogni componente

Considera il costo complessivo per l'organizzazione di ogni componente. Considera il costo totale di proprietà tenendo conto dei costi operativi e di gestione, soprattutto quando si utilizzano i servizi gestiti del provider cloud. L'attività di revisione deve riflettere i potenziali benefici (ad esempio, il tempo speso per l'analisi dovrebbe essere proporzionale al costo dei componenti).

Livello di rischio associato se questa best practice non fosse adottata: elevato

Guida all'implementazione

Considera il tempo risparmiato, che consentirà al proprio team di concentrarsi sul ritirare il debito tecnico, sull'innovazione e sulle funzionalità che offrono un valore aggiunto e sullo sviluppo di ciò che diversifica il business. Ad esempio, si potrebbe avere la necessità di eseguire il rehosting (lift and shift) del proprio database dall'ambiente on-premises nel cloud il più rapidamente possibile ed eseguire l'ottimizzazione in un secondo momento. Vale la pena soffermarsi sul risparmio possibile che puoi ottenere usando i servizi gestiti su AWS che rimuovono o riducono i costi di licenza. Servizi gestiti che AWS eliminano gli oneri operativi e amministrativi legati alla manutenzione di un servizio, come l'applicazione di patch o l'aggiornamento del sistema operativo, e consentono di concentrarsi sull'innovazione e sul business.

Dato che i servizi gestiti operano su scala cloud, possono offrire un costo inferiore per transazione o servizio. Questo vuol dire fare alcune ottimizzazioni potenziali in modo da ottenere benefici tangibili, senza modificare l'architettura principale dell'applicazione. Ad esempio, potresti voler ridurre la quantità di tempo dedicato alla gestione delle istanze di database migrando a una database-as-a-service piattaforma come Amazon Relational Database Service (AmazonRDS) o migrando la tua applicazione su una piattaforma completamente gestita come. AWS Elastic Beanstalk

Solitamente, i servizi gestiti presentano attributi che si possono impostare per garantire la capacità necessaria. Devi impostare e monitorare questi attributi in modo che la tua capacità in eccesso sia mantenuta al minimo e le prestazioni siano massimizzate. Puoi modificare gli attributi di AWS Managed Services utilizzo della sala operatoria AWS APIs e SDKs allineare le AWS Management Console esigenze di risorse all'evoluzione della domanda. Ad esempio, puoi aumentare o diminuire il numero di nodi su un EMR cluster Amazon (o su un cluster Amazon Redshift) per scalare orizzontalmente o internamente.

Puoi anche raggruppare più istanze su una AWS risorsa per attivare un utilizzo a maggiore densità. Ad esempio, puoi effettuare il provisioning di più piccoli database su una singola istanza di database Amazon Relational Database Service (RDSAmazon). Man mano che l'utilizzo aumenta, puoi migrare

uno dei database su un'istanza di RDS database Amazon dedicata utilizzando uno snapshot e un processo di ripristino.

Quando predisponi carichi di lavoro su servizi gestiti, devi comprendere i requisiti inerenti alla modifica della capacità del servizio. Tali requisiti solitamente riguardano il tempo, l'impegno e qualunque impatto sul normale funzionamento del carico di lavoro. La risorsa allocata deve offrire il tempo necessario per l'applicazione delle modifiche, pertanto procurati i mezzi necessari a tal fine. Lo sforzo continuo richiesto per modificare i servizi può essere ridotto praticamente a zero utilizzando APIs e SDKs integrando strumenti di sistema e monitoraggio, come Amazon CloudWatch.

Amazon RDS, Amazon Redshift e Amazon ElastiCache forniscono un servizio di database gestito. Amazon AthenaEMR, Amazon e Amazon OpenSearch Service forniscono un servizio di analisi gestito.

AMSè un servizio che gestisce AWS l'infrastruttura per conto di clienti e partner aziendali. Fornisce un ambiente sicuro e conforme in cui è possibile distribuire i carichi di lavoro. AMSutilizza modelli operativi cloud aziendali con automazione per consentire di soddisfare i requisiti dell'organizzazione, passare al cloud più rapidamente e ridurre i costi di gestione continui.

Passaggi dell'implementazione

- Esegui un'analisi completa: utilizzando l'elenco dei componenti, analizza ogni componente dalla priorità più alta alla priorità più bassa. Per la priorità più alta e i componenti più costosi, esegui analisi aggiuntive e valuta tutte le opzioni disponibili e il loro impatto a lungo termine. Per i componenti con priorità più bassa, valuta se le modifiche relative all'utilizzo hanno un impatto sulla priorità del componente, quindi esegui un'analisi dello sforzo appropriato.
- Confronta risorse gestite e non gestite: considera i costi operativi delle risorse gestite e confrontali con le risorse AWS gestite. Ad esempio, esamina i tuoi database in esecuzione su EC2 istanze Amazon e confrontali con Amazon RDS options (un servizio AWS gestito) o Amazon EMR rispetto all'esecuzione di Apache Spark su Amazon. EC2 Quando passi da un carico di lavoro autogestito a uno AWS completamente gestito, valuta attentamente le opzioni. I tre fattori più importanti da prendere in considerazione sono il tipo di servizio gestito da utilizzare, il processo che utilizzerai per la migrazione dei dati e la conoscenza del modello di responsabilità condivisa AWS.

Risorse

Documenti correlati:

- AWS Calcolatore del costo totale di proprietà () TCO

- Classi di archiviazione Amazon S3
- Prodotti Cloud AWS
- AWS Modello di responsabilità condivisa

Video correlati:

- Why move to a managed database?
- Cos'è Amazon EMR e come posso utilizzarlo per l'elaborazione dei dati?

Esempi correlati:

- Perché passare a un database gestito
- Consolida i dati da database SQL Server identici in un unico database Amazon RDS for SQL Server utilizzando AWS DMS
- Distribuisci dati su larga scala ad Amazon Managed Streaming for Apache Kafka (Amazon) MSK
- Migra un. ASP NETapplicazione web a AWS Elastic Beanstalk

COST05-BP04 Selezione del software con licenze convenienti

Il software open source elimina i costi di licenza del software, che contribuiscono in modo significativo ai costi dei carichi di lavoro. Laddove è richiesto un software concesso in licenza, evitate le licenze legate ad attributi arbitrariCPUs, ad esempio cercate licenze legate all'output o ai risultati. Il costo di queste licenze si ridimensiona in base ai vantaggi che offrono.

Livello di rischio associato se questa best practice non fosse adottata: basso

Guida all'implementazione

Il concetto di open source è nato nel contesto dello sviluppo del software per indicare che il software è conforme a determinati criteri di distribuzione gratuita. Il software open source è composto da codice sorgente che chiunque può analizzare, modificare e migliorare. In base ai requisiti aziendali, alle competenze degli ingegneri, all'utilizzo previsto o ad altre dipendenze tecnologiche, le organizzazioni possono prendere in considerazione l'utilizzo di software open source per ridurre al minimo i costi delle licenze. AWS In altri termini, utilizzando software open source è possibile ridurre il costo delle licenze software. Con lo scalare del carico di lavoro, l'impatto sui costi può essere significativo.

Misura i vantaggi di usare software con licenza in rapporto ai costi totali per ottimizzare il carico di lavoro. Crea modelli per le eventuali modifiche alla licenza e il relativo impatto sui costi del carico di lavoro. Se un fornitore modifica il costo della licenza del database, valuta come questo incide sull'efficienza complessiva del carico di lavoro. Effettua un'analisi dello storico dei prezzi dei tuoi fornitori per scoprire le tendenze dei cambiamenti relativi alle licenze dei loro prodotti. I costi delle licenze possono inoltre variare indipendentemente dalla velocità effettiva o dall'utilizzo, ad esempio per le licenze scalabili in base all'hardware (licenze vincolate). CPU È necessario evitare queste licenze poiché i costi possono aumentare rapidamente senza che vi siano vantaggi correlati.

Ad esempio, l'utilizzo di un'EC2istanza Amazon in us-east-1 con un sistema operativo Linux consente di ridurre i costi di circa il 45% rispetto all'esecuzione di un'altra istanza EC2 Amazon su Windows.

AWS Pricing CalculatorOffre un modo completo per confrontare i costi di varie risorse con diverse opzioni di licenza, come RDS istanze Amazon e diversi motori di database. Inoltre, AWS Cost Explorer offre una prospettiva inestimabile per i costi dei carichi di lavoro esistenti, in particolare quelli forniti con licenze diverse. Per la gestione delle licenze, AWS License Manager offre una soluzione semplificata per supervisionare e gestire le licenze software. I clienti possono implementare e rendere operativo il loro software open source preferito nel Cloud AWS.

Passaggi dell'implementazione

- Analizza le opzioni di licenza: esamina i termini di licenza del software disponibile. Cerca le versioni open source che dispongono delle funzionalità necessarie e considera se i vantaggi del software con licenza superano i costi. Condizioni convenienti possono rendere il costo del software proporzionato ai vantaggi che offre.
- Analizza i fornitori del software: esamina tutte le modifiche ai prezzi o alle licenze effettuate dal fornitore. Identifica eventuali modifiche non allineate ai risultati, ad esempio termini punitivi per l'esecuzione su hardware o piattaforme di fornitori specifici. Inoltre, verifica il modo in cui vengono eseguiti gli audit e le sanzioni in cui potresti incorrere.

Risorse

Documenti correlati:

- Open Source presso AWS
- AWS Calcolatore del costo totale di proprietà (TCO)
- Classi di archiviazione Amazon S3
- Prodotti cloud

Esempi correlati:

- Blog sull'open source
- AWS Blog open source
- Optimization and Licensing Assessment

COST05-BP05 Seleziona i componenti di questo carico di lavoro per ottimizzare i costi in linea con le priorità dell'organizzazione

Tieni in considerazione il costo nella selezione di tutti i componenti del tuo carico di lavoro. Ciò include l'utilizzo di servizi a livello di applicazione e servizi gestiti o serverless, container o un'architettura basata sugli eventi per ridurre i costi complessivi. Riduci al minimo i costi di licenza utilizzando software open source, software che non hanno costi di licenza o altre alternative per contenere la spesa.

Livello di rischio associato se questa best practice non fosse adottata: medio

Guida all'implementazione

Quando si selezionano tutti i componenti, è necessario considerare il costo dei servizi e delle opzioni. Ciò include l'utilizzo di servizi gestiti e a livello di applicazione, come Amazon Relational Database Service (Amazon), RDS Amazon DynamoDB, Amazon Simple Notification Service (Amazon) e SNS Amazon Simple Email Service (SESAmazon) per ridurre i costi complessivi dell'organizzazione.

Utilizza funzioni serverless e container per il calcolo, come AWS Lambda e Amazon Simple Storage Service (Amazon S3) per i siti Web statici. Se possibile, containerizza la tua applicazione e utilizza AWS Managed Container Services come Amazon Elastic Container Service (AmazonECS) o Amazon Elastic Kubernetes Service (Amazon). EKS

Riduci al minimo i costi di licenza utilizzando software open source o software che non prevedono tariffe di licenza: ad esempio, Amazon Linux per carichi di lavoro di calcolo oppure esegui la migrazione dei database ad Amazon Aurora.

Puoi utilizzare servizi serverless o a livello di applicazione come Lambda, Amazon Simple Queue Service (Amazon)SQS, Amazon e Amazon. SNS SES Questi servizi eliminano la necessità di gestire una risorsa e forniscono funzioni di esecuzione del codice, servizi di accodamento e consegna dei messaggi. L'altro vantaggio consiste nel ridurre orizzontalmente le prestazioni e i costi in base all'utilizzo, garantendo l'allocazione e l'attribuzione dei costi in modo efficiente.

L'utilizzo dell'architettura basata sugli eventi è inoltre possibile con i servizi serverless. Le architetture basate su eventi funzionano su base push, per cui tutto succede on demand quando l'evento si presenta sul router. In questo modo non devi sostenere i costi di un continuo polling per verificare un evento. Ciò significa un minore consumo di larghezza di banda di rete, un minore CPU utilizzo, una minore capacità del parco veicoli inattivo e un minor numero di strette di mano. SSL TLS

Per ulteriori informazioni sulle funzioni serverless, consulta il whitepaper Well-Architected Serverless Application lens.

Passaggi dell'implementazione

- Seleziona ciascun servizio per ottimizzare i costi: utilizzando l'elenco e l'analisi prioritari, seleziona ciascuna opzione che fornisce la migliore corrispondenza con le priorità dell'organizzazione. Invece di aumentare la capacità per soddisfare la domanda, prendi in considerazione altre opzioni che potrebbero offrirti performance migliori a costi inferiori. Ad esempio, se devi esaminare il traffico previsto per i tuoi database AWS, valuta la possibilità di aumentare le dimensioni dell'istanza o utilizzare ElastiCache i servizi Amazon (Redis o Memcached) per fornire meccanismi di memorizzazione nella cache per i tuoi database.
- Valuta l'architettura basata sugli eventi: l'utilizzo dell'architettura serverless consente inoltre di costruire un'architettura basata sugli eventi per applicazioni distribuite basate su microservizi, che aiuta a costruire soluzioni scalabili, resilienti, agili ed economiche.

Risorse

Documenti correlati:

- AWS Calcolatore del costo totale di proprietà () TCO
- Serverless in AWS
- Che cos'è un'architettura basata sugli eventi (EDA)?
- Classi di archiviazione Amazon S3
- Prodotti cloud
- Amazon ElastiCache (RedisOSS)

Esempi correlati:

- Getting started with event-driven architecture
- Architettura basata su eventi

- In che modo Statsig esegue Statsig in modo 100 volte più conveniente utilizzando Amazon (Redis) ElastiCache OSS
- Le migliori pratiche per lavorare con le funzioni AWS Lambda

COST05-BP06 Eseguire l'analisi dei costi per usi diversi nel tempo

I carichi di lavoro possono cambiare nel corso del tempo. Alcuni servizi o funzionalità sono più convenienti a diversi livelli di utilizzo. Eseguendo l'analisi su ogni componente nel tempo e in base all'utilizzo previsto, il carico di lavoro rimane conveniente per tutta la sua durata.

Livello di rischio associato se questa best practice non fosse adottata: medio

Guida all'implementazione

Con il AWS rilascio di nuovi servizi e funzionalità, i servizi ottimali per il carico di lavoro potrebbero cambiare. Tale cambiamento comporta un impegno, che dovrebbe essere commensurato ai vantaggi potenziali. La frequenza di revisione del carico di lavoro dipende dai requisiti dell'organizzazione. Se si tratta di un carico di lavoro con costi importanti, una rapida implementazione dei nuovi servizi massimizzerà i risparmi sui costi. In tal caso una revisione più frequente può risultare vantaggiosa. Un altro stimolo importante per la revisione è il cambiamento dei modelli di utilizzo. Se si verificassero notevoli cambiamenti nell'utilizzo, ciò potrebbe indicare un maggiore vantaggio dei servizi alternativi.

Se hai bisogno di trasferire dati Cloud AWS, puoi selezionare un'ampia varietà di servizi AWS offerti e strumenti partner per aiutarti a migrare i tuoi set di dati, che si tratti di file, database, immagini di macchine, volumi a blocchi o persino backup su nastro. Ad esempio, per spostare una grande quantità di dati da e verso AWS o elaborarli all'edge, puoi utilizzare uno dei dispositivi AWS appositamente progettati per spostare in modalità offline petabyte di dati in modo conveniente. Un altro esempio è rappresentato da velocità di trasferimento dati più elevate, un servizio di connessione diretta può essere più economico di un servizio VPN che fornisce la connettività costante richiesta per l'azienda.

In base all'analisi dei costi per usi diversi nel tempo, rivedi le tue attività di dimensionamento. Analizza i risultati per vedere se la policy di dimensionamento può essere ottimizzata per aggiungere istanze con tipi di istanze e opzioni di acquisto diversi. Esamina le tue impostazioni per vedere se il minimo può essere ridotto per soddisfare le richieste degli utenti, ma con una dimensione inferiore del parco istanze, e aggiungi più risorse per i momenti attesi di incremento della domanda.

Esegui analisi dei costi per i vari utilizzi nel tempo discutendo con le parti interessate della tua organizzazione e utilizza la funzionalità di previsione di AWS Cost Explorer per anticipare il potenziale

impatto delle modifiche ai servizi. Monitora i livelli di utilizzo Budget AWS, utilizza gli allarmi di CloudWatch fatturazione e identifica e AWS Cost Anomaly Detection implementa prima i servizi più convenienti.

Passaggi dell'implementazione

- Definisci modelli di utilizzo previsti: collaborando con la tua organizzazione, ad esempio con i proprietari di prodotti e marketing, documenta quali sono i modelli di utilizzo previsti per il carico di lavoro. Discuti con le parti interessate dell'azienda dell'aumento dell'utilizzo e dei costi storici e previsti e verifica che tali incrementi siano in linea con i requisiti aziendali. Identifica i giorni, le settimane o i mesi di calendario in cui prevedi che più utenti utilizzino AWS le tue risorse, il che indica che dovresti aumentare la capacità delle risorse esistenti o adottare servizi aggiuntivi per ridurre i costi e aumentare le prestazioni.
- Esegui l'analisi dei costi in base all'utilizzo previsto: esegui l'analisi in ciascuno di questi punti mediante i modelli di utilizzo definiti. Lo sforzo di analisi dovrebbe riflettere il potenziale risultato. Ad esempio, se la variazione dell'utilizzo è elevata, è necessario eseguire un'analisi accurata per verificare eventuali costi e modifiche. In altre parole, quando il costo aumenta dovrebbe aumentare anche l'utilizzo per l'azienda.

Risorse

Documenti correlati:

- AWS Calcolatore del costo totale di proprietà (TCO)
- Classi di archiviazione Amazon S3
- Prodotti cloud
- Amazon EC2 Auto Scaling
- Migrazione dei dati nel cloud
- AWS Snow Family

Video correlati:

- AWS OpsHub for Snow Family

COST6. In che modo raggiungi gli obiettivi di costo quando selezioni il tipo, le dimensioni e il numero delle risorse?

Assicurati di scegliere la dimensione e il numero delle risorse appropriati per l'attività in questione. Selezionando il tipo, le dimensioni e il numero più convenienti, riduci al minimo gli sprechi.

Best practice

- COST06-BP01 Eseguire la modellazione dei costi
- COST06-BP02 Seleziona il tipo, la dimensione e il numero di risorse in base ai dati
- COST06-BP03 Seleziona automaticamente il tipo, la dimensione e il numero di risorse in base alle metriche
- COST06-BP04 Prendi in considerazione l'utilizzo di risorse condivise

COST06-BP01 Eseguire la modellazione dei costi

Identifica i requisiti dell'organizzazione (come le esigenze aziendali e gli impegni esistenti) ed esegui la modellazione dei costi (costi complessivi) del carico di lavoro e di ciascuno dei suoi componenti. Esegui benchmark per il carico di lavoro in base ai diversi carichi previsti e confronta i costi. L'impegno di modellazione deve riflettere il potenziale risultato. Ad esempio, il tempo speso è proporzionale al costo dei componenti.

Livello di rischio associato se questa best practice non fosse adottata: elevato

Guida all'implementazione

Esegui la modellazione dei costi per il tuo carico di lavoro e ciascuno dei suoi componenti per stabilire il giusto equilibrio tra le risorse e trova la dimensione appropriata per ciascuna risorsa nel carico di lavoro, sulla base di un determinato livello di prestazioni. La comprensione delle considerazioni sui costi può contribuire a business case dell'organizzazione e processo decisionale quando si valutano i risultati di realizzazione del valore per l'implementazione del carico di lavoro pianificato.

Esegui benchmark per il carico di lavoro in base ai diversi carichi previsti e confronta i costi. L'impegno di modellazione deve riflettere il potenziale risultato. Ad esempio, il tempo speso dovrebbe essere proporzionale al costo dei componenti o ai risparmi previsti. Per le best practice, consulta la sezione Review del Performance Efficiency Pillar del AWS Well-Architected Framework.

Ad esempio, per la creazione di modelli dei costi per un carico di lavoro costituito da risorse di calcolo, AWS Compute Optimizer può contribuire alla modellazione dei costi per l'esecuzione dei

carichi di lavoro. Fornisce consigli di dimensionamento appropriato per le risorse di calcolo in base a una valutazione cronologica dell'utilizzo. Assicurati che CloudWatch gli agenti siano distribuiti EC2 sulle istanze Amazon per raccogliere metriche di memoria che ti aiutino a fornire consigli più accurati all'interno. AWS Compute Optimizer Questa è l'origine dati ideale per le risorse di calcolo poiché si tratta di un servizio gratuito e utilizza il machine learning per generare più raccomandazioni a seconda dei livelli di rischio.

Esistono diversi servizi che puoi utilizzare con i log personalizzati come fonti di dati per operazioni di dimensionamento corretto per altri servizi e componenti del carico di lavoro, come Amazon e AWS Trusted AdvisorAmazon Logs. CloudWatch CloudWatch AWS Trusted Advisor controlla le risorse e contrassegna le risorse a basso utilizzo, il che può aiutarti a dimensionare correttamente le risorse e creare modelli di costo.

Di seguito sono riportate le raccomandazioni relative a parametri e dati di modellazione dei costi:

- Il monitoraggio deve corrispondere in modo preciso all'esperienza degli utenti. Seleziona la granularità corretta per un dato periodo di tempo e scegli in modo ponderato il 99° percentile o quello massimo invece del valore medio.
- Seleziona la granularità corretta per il periodo di analisi richiesto per coprire tutti i cicli del carico di lavoro. Ad esempio, se esegui un'analisi di due settimane, potresti ignorare un ciclo mensile di utilizzo elevato, con conseguente provisioning insufficiente.
- Scegliete i AWS servizi giusti per il carico di lavoro pianificato tenendo conto degli impegni esistenti, dei modelli di prezzo selezionati per altri carichi di lavoro e della capacità di innovare più rapidamente e concentrarvi sul valore aziendale principale.

Passaggi dell'implementazione

- Esegui la modellazione dei costi per le risorse: implementa il carico di lavoro o una proof of concept in un account separato con i tipi di risorse e le dimensioni specifiche da testare. Esegui il carico di lavoro con i dati di test e registra i risultati di output, insieme ai dati relativi ai costi per il tempo in cui è stato eseguito il test. Quindi, implementa di nuovo il carico di lavoro o modifica tipi e dimensioni delle risorse ed esegui nuovamente il test. Includi i costi di licenza di qualsiasi prodotto che si possa utilizzare con queste risorse e i costi operativi stimati (manodopera o tecnici) per l'implementazione e la gestione di queste risorse durante la creazione di modelli di costo. Considera la modellazione dei costi per un periodo (orario, giornaliero, mensile, annuale o triennale).

Risorse

Documenti correlati:

- AWS Auto Scaling
- Identificare le opportunità per il ridimensionamento corretto
- CloudWatch Funzionalità di Amazon
- Ottimizzazione dei costi: Amazon EC2 Right Sizing
- AWS Compute Optimizer
- AWS Calcolatore dei prezzi

Esempi correlati:

- Esegui una modellazione dei costi basata sui dati
- Stima il costo delle configurazioni pianificate AWS delle risorse
- Scegli gli strumenti giusti AWS

COST06-BP02 Seleziona il tipo, la dimensione e il numero di risorse in base ai dati

Seleziona la dimensione o il tipo di risorsa in base ai dati relativi a carico di lavoro e caratteristiche delle risorse come, ad esempio, calcolo, memoria, throughput o scrittura intensiva. Questa selezione è tipicamente effettuata utilizzando una versione precedente (on-premises) del carico di lavoro, utilizzando la documentazione o altre fonti di informazione sul carico di lavoro.

Livello di rischio associato se questa best practice non fosse adottata: medio

Guida all'implementazione

Amazon EC2 offre un'ampia selezione di tipi di istanze con diversi livelli di memoriaCPU, storage e capacità di rete per adattarsi a diversi casi d'uso. Questi tipi di istanze offrono diverse combinazioni di funzionalità di memoriaCPU, archiviazione e rete, che ti offrono versatilità nella scelta della giusta combinazione di risorse per i tuoi progetti. Ogni tipo di istanza è disponibile in più dimensioni, per consentire di adattare le risorse alle richieste del carico di lavoro. Per determinare il tipo di istanza necessario, acquisisci i dettagli sui requisiti di sistema dell'applicazione o del software che intendi eseguire sull'istanza. Tali dettagli devono includere le informazioni seguenti:

- Sistema operativo
- Numero di core CPU

- GPUnuclei
- Quantità di memoria di sistema () RAM
- Tipo e spazio di archiviazione
- Requisiti di larghezza di banda della rete

Identifica lo scopo dei requisiti di calcolo e l'istanza necessaria, quindi esplora le varie famiglie di EC2 istanze Amazon. Amazon offre le seguenti famiglie di tipi di istanza:

- Uso generico
- Ottimizzata per il calcolo
- Ottimizzata per la memoria
- Storage ottimizzato
- Calcolo accelerato
- HPCOttimizzato

Per una comprensione più approfondita degli scopi e dei casi d'uso specifici che una particolare famiglia di EC2 istanze Amazon può soddisfare, consulta Tipi di AWS istanze.

L'acquisizione dei requisiti di sistema è fondamentale per selezionare famiglia e tipo di istanze specifici più adatti alle proprie esigenze. I nomi dei tipi di istanza sono composti dal nome della famiglia e dalla dimensione dell'istanza. Ad esempio, l'istanza t2.micro appartiene alla famiglia T2 ed è di dimensioni ridotte.

Seleziona la dimensione o il tipo di risorsa in base al carico di lavoro e alle caratteristiche delle risorse come, ad esempio, calcolo, memoria, throughput o uso intensivo di operazioni di scrittura. Questa selezione è in genere effettuata ricorrendo alla modellazione dei costi, a una versione precedente del carico di lavoro (ad esempio una versione on-premises), alla documentazione o ad altre fonti di informazione sul carico di lavoro (come whitepaper o soluzioni pubblicate). L'utilizzo di calcolatori AWS dei prezzi o strumenti di gestione dei costi può aiutare a prendere decisioni informate su tipi, dimensioni e configurazioni delle istanze.

Passaggi dell'implementazione

- Seleziona le risorse in base ai dati: usa i dati sulla modellazione del costo per selezionare il livello di utilizzo del carico di lavoro previsto e scegli il tipo e le dimensioni delle risorse specificate. Basandoti sui dati di modellazione dei costi, determina il numero di memoria virtuale totale

(GiB)CPUs, il volume di archiviazione dell'istanza locale (GB), i EBS volumi Amazon e il livello di prestazioni della rete, tenendo conto della velocità di trasferimento dei dati richiesta per l'istanza. Effettua sempre selezioni basate su analisi dettagliate e dati accurati per ottimizzare le prestazioni e contemporaneamente gestire i costi in modo efficace.

Risorse

Documenti correlati:

- AWS Tipi di istanza
- AWS Auto Scaling
- CloudWatch Funzionalità di Amazon
- Ottimizzazione dei costi: EC2 giusto dimensionamento

Video correlati:

- Selezione dell'EC2istanza Amazon giusta per i tuoi carichi di lavoro
- Right size your service

Esempi correlati:

- Ora è diventato più facile scoprire e confrontare i tipi di EC2 istanze Amazon

COST06-BP03 Seleziona automaticamente il tipo, la dimensione e il numero di risorse in base alle metriche

Utilizza i parametri del carico di lavoro in esecuzione per selezionare la dimensione e il tipo corretti per ottimizzare i costi. Esegui il provisioning in modo corretto di throughput, dimensione e spazio di archiviazione per servizi di calcolo, memorizzazione, gestione dati e di rete. Questa operazione può essere eseguita con un ciclo di feedback, ad esempio il dimensionamento automatico o tramite codice personalizzato nel carico di lavoro.

Livello di rischio associato se questa best practice non fosse adottata: basso

Guida all'implementazione

Crea un ciclo di feedback all'interno del carico di lavoro che utilizza i parametri attivi del carico di lavoro in esecuzione per apportarvi modifiche. È possibile utilizzare un servizio gestito, ad esempio

configurato dall'utente per eseguire le AWS Auto Scalingoperazioni di dimensionamento corrette per le proprie esigenze. AWS fornisce APIs inoltre funzionalità che consentono di modificare le risorse con il minimo sforzo. SDKs Puoi programmare un carico di lavoro su stop-and-start un'EC2istanza Amazon per consentire la modifica delle dimensioni o del tipo di istanza. Ciò offre i vantaggi del dimensionamento appropriato, eliminando al contempo quasi tutti i costi operativi necessari per apportare la modifica.

Alcuni AWS servizi dispongono di una selezione automatica del tipo o della dimensione, come Amazon Simple Storage Service Intelligent-Tiering. Il Piano intelligente Amazon S3 sposta automaticamente i dati tra due livelli di accesso: frequente e poco frequente, in base ai tuoi modelli di utilizzo.

Passaggi dell'implementazione

- Aumenta l'osservabilità configurando le metriche del carico di lavoro: acquisisci le metriche chiave per il carico di lavoro. Queste metriche forniscono un'indicazione dell'esperienza del cliente, ad esempio l'output del carico di lavoro, e si allineano alle differenze tra tipi e dimensioni delle risorse, ad esempio l'utilizzo della memoria. CPU Per quanto riguarda le risorse di calcolo, analizza i dati sulle prestazioni per dimensionare correttamente le tue EC2 istanze Amazon. Identifica le istanze inattive e quelle sottoutilizzate. Le metriche chiave da cercare sono CPU l'utilizzo e l'utilizzo della memoria (ad esempio, CPU utilizzo del 40% nel 90% delle volte, come spiegato in Rightsizing with and Memory Utilization Enabled). AWS Compute Optimizer Identifica le istanze con un CPU utilizzo massimo e un utilizzo della memoria inferiori al 40% in un periodo di quattro settimane. Queste sono le istanze in cui occorre dimensionare correttamente il sistema per ridurre i costi. Per le risorse di archiviazione, come Amazon S3, puoi utilizzare Amazon S3 Storage Lens, che per impostazione predefinita ti consente di visualizzare 28 parametri in varie categorie a livello di bucket e 14 giorni di dati storici nei pannelli di controllo. Per analizzare specifici parametri, si possono applicare dei filtri su riepilogo e ottimizzazione dei costi o eventi all'interno del pannello di controllo di Amazon S3 Storage Lens.
- Visualizza i consigli per il corretto dimensionamento: utilizza i consigli sul corretto dimensionamento in e lo strumento Amazon EC2 rightsizing AWS Compute Optimizer nella console di gestione dei costi, oppure verifica il corretto AWS Trusted Advisor dimensionamento delle risorse per apportare modifiche al carico di lavoro. È importante utilizzare gli strumenti giusti per dimensionare correttamente le diverse risorse e seguire le linee guida per il corretto dimensionamento, indipendentemente dal fatto che si tratti di un'EC2istanza Amazon, di classi di AWS storage o di tipi di istanze Amazon. RDS Per le risorse di archiviazione è possibile utilizzare Amazon S3 Storage Lens, che offre visibilità sull'utilizzo dello spazio di archiviazione di oggetti e sulle tendenze delle attività e fornisce raccomandazioni operative per ottimizzare i costi e applicare le

best practice di protezione dei dati. Grazie ai consigli contestuali che Amazon S3 Storage Lens estrapola dall'analisi dei parametri all'interno dell'organizzazione, puoi adottare misure immediate per ottimizzare l'archiviazione.

- Seleziona il tipo di risorse ed esegui il dimensionamento in automatico sulla base delle metriche: utilizza i parametri del carico di lavoro per seleziona manualmente o in automatico le risorse del carico di lavoro. Per le risorse di calcolo, la configurazione di AWS Auto Scaling o l'implementazione di codice all'interno dell'applicazione può ridurre lo sforzo necessario in caso di modifiche frequenti e permettere di implementare potenzialmente eventuali modifiche più velocemente rispetto a un processo manuale. È possibile avviare e scalare automaticamente un parco di istanze on demand e istanze spot all'interno di un singolo gruppo con un singolo gruppo Auto Scaling. Oltre a ricevere sconti per l'utilizzo di Istanze Spot, è possibile utilizzare Istanze riservate o Savings Plan per ricevere tariffe scontate sul normale prezzo delle istanze on demand. Tutti questi fattori combinati ti aiutano a ottimizzare i risparmi sui costi per EC2 le istanze Amazon e a determinare la scalabilità e le prestazioni desiderate per la tua applicazione. Puoi anche utilizzare una strategia di selezione del tipo di istanza (ABS) basata sugli attributi in Auto Scaling Groups (ASG), che ti consente di esprimere i requisiti dell'istanza come un insieme di attributi, come vCPU, memoria e archiviazione. Puoi utilizzare automaticamente i tipi di istanze di nuova generazione quando vengono rilasciati e accedere a una gamma più ampia di capacità con Amazon EC2 Spot Instances. Amazon EC2 Fleet e Amazon EC2 Auto Scaling selezionano e avviano le istanze che soddisfano gli attributi specificati, eliminando la necessità di selezionare manualmente i tipi di istanza. Per le risorse di storage, puoi utilizzare le funzionalità di Amazon S3 Intelligent Tiering EFS e Amazon Infrequent Access, che consentono di selezionare automaticamente classi di storage che offrono risparmi automatici sui costi di storage quando cambiano i modelli di accesso ai dati, senza impatto sulle prestazioni o sovraccarico operativo.

Risorse

Documenti correlati:

- AWS Auto Scaling
- AWS Dimensionamento corretto
- AWS Compute Optimizer
- CloudWatch Funzionalità di Amazon
- CloudWatchConfigurazione iniziale
- CloudWatchPubblicazione di metriche personalizzate
- Guida introduttiva ad Amazon EC2 Auto Scaling

- Amazon S3 Storage Lens
- Amazon S3 Intelligent-Tiering
- Accesso EFS non frequente ad Amazon
- Avvia un'EC2istanza Amazon utilizzando il SDK

Video correlati:

- Right Size Your Services

Esempi correlati:

- Selezione del tipo di istanza basata sugli attributi per Auto Scaling for Amazon Fleet EC2
- Optimizing Amazon Elastic Container Service for cost using scheduled scaling
- Scalabilità predittiva con Amazon EC2 Auto Scaling
- Optimize Costs and Gain Visibility into Usage with Amazon S3 Storage Lens
- Well-Architected Labs: raccomandazioni per il ridimensionamento corretto (Livello 100)

COST06-BP04 Prendi in considerazione l'utilizzo di risorse condivise

Per i servizi già distribuiti a livello di organizzazione per più unità aziendali, prendete in considerazione l'utilizzo di risorse condivise per aumentare l'utilizzo e ridurre il costo totale di proprietà (). TCO L'utilizzo delle risorse condivise può essere un'opzione conveniente per centralizzare gestione e costi mediante le soluzioni esistenti, condividendo i componenti o in entrambi i casi. Gestisci le funzioni comuni, come monitoraggio, backup e connettività, entro il limite dell'account o in un account dedicato. Inoltre, puoi diminuire i costi implementando la standardizzazione, riducendo duplicazione e complessità.

Livello di rischio associato se questa best practice non fosse adottata: medio

Guida all'implementazione

Se più carichi di lavoro eseguono la stessa funzione, utilizza soluzioni esistenti e componenti condivisi per migliorare la gestione e ottimizzare i costi. Prendi in considerazione l'utilizzo delle risorse esistenti, in particolare quelle condivise, come server di database non di produzione o servizi di directory, per contenere i costi del cloud seguendo le best practice di sicurezza e le normative dell'organizzazione. Per realizzare valore ed efficienza ottimali, è fondamentale utilizzare report di

showback e meccanismi di chargeback per riallocare i costi alle aree pertinenti dell'azienda che determinano i consumi.

Con showback si fa riferimento ai report che suddividono i costi del cloud in categorie attribuibili, come consumatori, business unit, account di contabilità generale o altre entità responsabili. L'obiettivo dei report di showback è mostrare a team, business unit o singole persone il costo delle risorse cloud consumate.

Per chargeback si intende l'allocazione della spesa per i servizi centrali alle unità di costo in base a una strategia adatta a uno specifico processo di gestione finanziaria. Per i clienti, il chargeback addebita il costo sostenuto da un account di servizi condivisi a diverse categorie di costi finanziari definite per un processo di report dei clienti. Stabilendo i meccanismi di chargeback, puoi dichiarare i costi sostenuti da diverse business unit, prodotti e team.

È possibile classificare i carichi di lavoro come critici e non critici. Sulla base di tale classificazione, utilizza le risorse condivise con configurazioni generali per i carichi di lavoro meno critici. Per ottimizzare ulteriormente i costi, usa i server dedicati esclusivamente per i carichi di lavoro critici. Condividi o alloca le risorse in più account per gestirle in modo efficiente. La condivisione è sicura e non compromette la struttura organizzativa anche in caso di separazione di ambienti di sviluppo, test e produzione.

Per migliorare la comprensione e ottimizzare i costi e l'utilizzo delle applicazioni containerizzate, utilizza i dati di allocazione dei costi suddivisi che consentono di allocare i costi alle singole entità aziendali in base al modo in cui l'applicazione consuma le risorse di calcolo e memoria condivise. La suddivisione dei dati sull'allocazione dei costi ti aiuta a ottenere uno showback e un chargeback a livello di task nei carichi di lavoro dei container in esecuzione su Amazon Elastic Container Service (Amazon) o Amazon Elastic ECS Kubernetes Service (Amazon). EKS

Per le architetture distribuite, crea un servizio condiviso, che fornisca l'accesso centralizzato ai servizi condivisi richiesti dai carichi di VPC lavoro in ciascuna di esse. VPCs Questi servizi condivisi possono includere risorse come servizi di directory o endpoint. VPC Per ridurre il sovraccarico e i costi amministrativi, condividi le risorse da una postazione centrale anziché costruirle in ciascuna di esse. VPC

Quando si utilizzano le risorse condivise, è possibile ridurre i costi operativi, massimizzare l'utilizzo delle risorse e migliorare la coerenza. In una progettazione con più account, è possibile ospitare alcuni AWS servizi centralmente e accedervi utilizzando diverse applicazioni e account in un hub per risparmiare sui costi. Puoi usare AWS Resource Access Manager (AWS RAM) per condividere altre risorse comuni, come VPCsottoreti e AWS Transit Gateway allegati AWS Network Firewallo Amazon

pipeline. SageMaker In un ambiente con più account, usa AWS RAM per creare una risorsa una sola volta e condividerla con altri account.

Le organizzazioni devono applicare i tag in modo efficace ai costi condivisi e verificare che non vi siano parti significative dei costi senza tag o allocazione. Se non si allocano i costi condivisi in modo efficace e nessuno se ne assume la responsabilità della gestione, i costi condivisi del cloud possono aumentare vertiginosamente. È necessario sapere dove sostieni i costi a livello di risorse, carico di lavoro, team oppure organizzazione poiché queste informazioni migliorano la tua comprensione del valore fornito al livello applicabile rispetto ai risultati aziendali raggiunti. In definitiva, le organizzazioni ottengono il vantaggio del risparmio sui costi grazie alla condivisione dell'infrastruttura cloud. Incoraggia l'allocazione dei costi sulle risorse condivise del cloud per ottimizzare la spesa del cloud.

Passaggi dell'implementazione

- Valutazione dele risorse esistenti: esamina i carichi di lavoro esistenti che utilizzano servizi simili per il tuo carico di lavoro. A seconda dei componenti del carico di lavoro, considera le piattaforme esistenti, se la logica aziendale o i requisiti tecnici lo consentono.
- Utilizza la condivisione delle risorse AWS RAM e limita di conseguenza: utilizza AWS RAM per condividere le risorse con altri AWS account all'interno dell'organizzazione. Con la condivisione non dovrai duplicare le risorse in più account e riduci al minimo l'onere operativo della manutenzione delle risorse. Questo processo ti consente inoltre di condividere in modo sicuro le risorse create con i ruoli e gli utenti del proprio account e di altri Account AWS.
- Tag delle risorse: effettua il tag delle risorse candidate alla rendicontazione dei costi e classificale in categorie di costo. Attiva questi tag di risorse relativi ai costi per l'allocazione dei costi per fornire visibilità sull'utilizzo delle AWS risorse. Concentrati sulla creazione di un livello di granularità appropriato per quanto riguarda la visibilità dei costi e dell'utilizzo e influenza i comportamenti di consumo del cloud attraverso il reporting e il monitoraggio dell'allocazione dei costi. KPI

Risorse

Best practice correlate:

- SEC03-BP08 Condividi le risorse in modo sicuro all'interno della tua organizzazione

Documenti correlati:

- Che cos' AWS Resource Access Managerè?
- AWS servizi con cui è possibile utilizzare AWS Organizations

- Risorse condivisibili AWS
- AWS Interrogazioni su costi e utilizzo (CUR)

Video correlati:

- AWS Resource Access Manager - controllo granulare degli accessi con autorizzazioni gestite
- Come progettare la strategia di allocazione dei AWS costi
- AWS Cost Categories

Esempi correlati:

- Come riaddebitare i servizi condivisi: un esempio AWS Transit Gateway
- Come creare un modello di chargeback/showback per Savings Plans utilizzando CUR
- Utilizzo della VPC condivisione per un'architettura di microservizi multi-account conveniente
- Migliora la visibilità dei costi di Amazon EKS con AWS Split Cost Allocation Data
- Migliora la visibilità dei costi di Amazon ECS e AWS Batch con AWS Split Cost Allocation Data

COST7. In che modo impieghi i modelli di prezzo per ridurre i costi?

Utilizza il modello di prezzo più appropriato per le tue risorse per ridurre al minimo le spese.

Best practice

- COST07-BP01 Eseguire l'analisi del modello di determinazione dei prezzi
- COST07-BP02 Scegli le regioni in base al costo
- COST07-BP03 Seleziona accordi di terze parti con condizioni convenienti
- COST07-BP04 Implementa modelli di prezzo per tutti i componenti di questo carico di lavoro
- COST07-BP05 Eseguire l'analisi del modello di determinazione dei prezzi a livello di account di gestione

COST07-BP01 Eseguire l'analisi del modello di determinazione dei prezzi

Analizza ogni componente del carico di lavoro. Determina se il componente e le risorse saranno in esecuzione per periodi prolungati (per sconti a fronte di impegni) o dinamici e di breve durata (per spot oppure on demand). Esegui un'analisi sul carico di lavoro utilizzando i suggerimenti degli

strumenti di gestione dei costi e applica le regole aziendali ai suggerimenti per ottenere rendimenti elevati.

Livello di rischio associato se questa best practice non fosse adottata: elevato

Guida all'implementazione

AWS dispone di diversi modelli di prezzo che consentono di pagare le risorse nel modo più conveniente, adatto alle esigenze dell'organizzazione e in base al prodotto. Lavora con i tuoi team per stabilire il modello di prezzi più appropriato. Spesso il modello di prezzi è costituito da più opzioni, in base alla tua disponibilità

Le istanze on demand consentono di pagare per la capacità di calcolo o di database all'ora o al secondo (minimo 60 secondi), in base alle istanze in esecuzione, senza impegni a lungo termine o pagamenti anticipati.

Savings Plans è un modello di prezzo flessibile che offre prezzi bassi su AmazonEC2, Lambda e AWS Fargate sull'utilizzo, in cambio dell'impegno a una quantità di utilizzo costante (misurata in dollari all'ora) per un periodo di uno o tre anni.

Le istanze Spot sono un meccanismo di EC2 determinazione dei prezzi di Amazon che consente di richiedere capacità di elaborazione di riserva a una tariffa oraria scontata (fino al 90% di sconto sul prezzo on demand) senza impegno anticipato.

Le istanze riservate offrono uno sconto fino al 75% pagando in anticipo la capacità. Per ulteriori informazioni, consulta Ottimizzazione dei costi con le prenotazioni.

Potresti scegliere di includere un Savings Plan per le risorse associate alla produzione, alla qualità e agli ambienti di sviluppo. In alternativa, poiché le risorse dell'ambiente di sperimentazione (sandbox) vengono attivate solo quando necessario, è possibile scegliere un modello on demand per le risorse di quell'ambiente. Usa Amazon Spot Instances per ridurre EC2 i costi di Amazon o utilizza Compute Savings Plans per ridurre i costi di AmazonEC2, Fargate e Lambda. Lo strumento per i suggerimenti di AWS Cost Explorer offre sconti sugli impegni con i Saving Plans.

Se hai acquistato istanze riservate per Amazon EC2 in passato o hai stabilito pratiche di allocazione dei costi all'interno della tua organizzazione, puoi continuare a utilizzare Amazon EC2 Reserved Instances per il momento. Tuttavia, ti consigliamo di lavorare su una strategia per usare i Savings Plans in futuro come meccanismo più flessibile di risparmio sui costi. Puoi aggiornare Savings Plans (SP) Recommendations AWS Cost Management per generare nuovi Savings Plans Recommendations in qualsiasi momento. Utilizza le istanze riservate (RI) per ridurre i costi di

AmazonRDS, Amazon Redshift, ElastiCache Amazon e OpenSearch Amazon Service. Savings Plans e le istanze riservate sono disponibili in tre opzioni: pagamento anticipato totale, pagamento anticipato parziale e nessun pagamento anticipato. Utilizza i consigli forniti nei consigli di acquisto AWS Cost Explorer RI e SP.

Per trovare opportunità per i carichi di lavoro Spot, utilizza una visualizzazione oraria dell'utilizzo complessivo e cerca periodi regolari di variazione di utilizzo o di elasticità. Puoi usare le istanze Spot per diverse applicazioni flessibili e con tolleranza ai guasti. Gli esempi includono server Web stateless, API endpoint, applicazioni di analisi e big data, carichi di lavoro containerizzati, CI/CD e altri carichi di lavoro flessibili.

Analizza le tue RDS istanze Amazon EC2 e Amazon per vedere se possono essere disattivate quando non le usi (fuori orario e nei fine settimana). In questo modo potrai ridurre i costi di almeno il 70% rispetto al loro utilizzo 24 ore su 24, 7 giorni su 7. Se hai cluster Amazon Redshift che devono essere disponibili solo in orari specifici, puoi metterli in pausa e poi riattivarli. Quando il cluster Amazon Redshift o Amazon and EC2 Amazon RDS Instance viene interrotto, la fatturazione di elaborazione si interrompe e viene applicata solo la tariffa di storage.

Tieni presente che le prenotazioni On-Demand Capacity (ODCR) non sono uno sconto sui prezzi. Le prenotazioni della capacità vengono addebitate alla tariffa on-demand equivalente indipendentemente dal fatto che si stia o meno eseguendo istanze nella capacità riservata. Tali prenotazioni devono essere prese in considerazione quando hai bisogno di offrire capacità sufficiente alle risorse che desideri eseguire. ODCRsnon devono essere vincolati a impegni a lungo termine, in quanto possono essere annullati quando non ne hai più bisogno, ma possono anche beneficiare degli sconti offerti da Savings Plans o Reserved Instances.

Passaggi dell'implementazione

- Analizza l'elasticità del carico di lavoro: utilizzando la granularità oraria in Cost Explorer o un pannello di controllo personalizzato, analizza l'elasticità del carico di lavoro. Vai alla ricerca di modifiche regolari del numero di istanze in esecuzione. Le istanze in esecuzione per brevi periodi di tempo sono candidate per essere istanze spot o parco istanze spot.
  - Well-Architected Lab: Cost Explorer
  - Well-Architected Labs: visualizzazione dei costi
- Esamina i contratti esistenti sui prezzi: esamina i contratti o gli impegni in essere per le esigenze a lungo termine. Analizza ciò di cui disponi ora e fino a che punto gli impegni presi vengono sfruttati. Sfrutta sconti contrattuali preesistenti o accordi aziendali. Gli accordi aziendali consentono ai clienti di personalizzare i contratti per adattarli alle loro esigenze. Per gli impegni a lungo termine, prendi

in considerazione sconti sui prezzi riservati, Reserved Instances o Savings Plans per il tipo di istanza, la famiglia di istanze e le zone di disponibilità specifici. Regione AWS

- Esegui un'analisi degli sconti a fronte di un impegno: con Cost Explorer nel tuo account, consulta i consigli relativi a Savings Plans e istanze riservate. Per verificare di implementare le raccomandazioni corrette con gli sconti e i rischi richiesti, segui i Well-Architected labs.

Risorse

Documenti correlati:

- Accessing Reserved Instance recommendations
- Opzioni di acquisto delle istanze
- AWS Impresa

Video correlati:

- Save up to 90% and run production workloads on Spot

Esempi correlati:

- Well-Architected Lab: Cost Explorer
- Well-Architected Labs: visualizzazione dei costi
- Well-Architected Lab: modelli di prezzo

COST07-BP02 Scegli le regioni in base al costo

La determinazione dei prezzi delle risorse può essere diversa in ciascuna regione. Individua le differenze di costo a livello regionale ed esegui la distribuzione solo nelle regioni con costi più elevati per soddisfare i requisiti di latenza, residenza dei dati e sovranità dei dati. La considerazione del costo della regione garantisce il pagamento del prezzo complessivo più basso per questo carico di lavoro.

Livello di rischio associato se questa best practice non fosse adottata: medio

Guida all'implementazione

L'Cloud AWS infrastruttura è globale, ospitata in più sedi in tutto il mondo e costruita attorno a Availability Zones Regioni AWS, Local Zones, AWS Outposts e Wavelength Zones. Una regione è un'ubicazione fisica nel mondo e ogni regione è un'area geografica separata in cui AWS sono presenti più zone di disponibilità. Le zone di disponibilità, che sono più sedi isolate all'interno di ogni regione, sono costituite da uno o più data center discreti, ciascuno con alimentazione, rete e connettività ridondanti.

Ciascuna Regione AWS opera secondo le condizioni del mercato locale e i prezzi delle risorse sono diversi in ogni regione a causa delle differenze nel costo del terreno, della fibra, dell'elettricità e delle tasse, ad esempio. Scegli una regione specifica per gestire un componente o tutta la tua soluzione in modo da eseguirla al minor prezzo possibile a livello globale. Usa il calcolatore AWS per stimare i costi del carico di lavoro in varie regioni, cercando i servizi per tipo di località (regione, zona di lunghezza d'onda e zona locale) e regione.

Quando progetti le tue soluzioni, una best practice da seguire è quella di cercare di posizionare le risorse di calcolo vicino agli utenti per offrire una latenza inferiore e una forte sovranità dei dati. Seleziona la posizione geografica in base alle esigenze di business, privacy dei dati, performance e requisiti di sicurezza. Per le applicazioni con utenti finali globali, utilizza più sedi.

Utilizza le regioni che offrono AWS servizi a prezzi più bassi per distribuire i tuoi carichi di lavoro se non hai obblighi in materia di privacy dei dati, sicurezza e requisiti aziendali. Ad esempio, se la tua regione predefinita è Asia Pacifico (Sydney) (`ap-southwest-2`) e se non ci sono restrizioni (privacy dei dati, sicurezza, ad esempio) all'utilizzo di altre regioni, la distribuzione di EC2 istanze Amazon non critiche (sviluppo e test) negli Stati Uniti orientali (Virginia settentrionale) (`us-east-1`) ti costerà meno.

	Conformità	Latenza	Costo	Servizi/Caratteristiche
Regione 1	✓	15 ms	$$	✓
Regione 2	✓	20 ms	$$$	X
Regione 3	✓	80 ms	$	✓
Regione 4	✓	15 ms	$$	✓
Regione 5	✓	20 ms	$$$	X
Regione 6	✓	15 ms	$	✓
Regione 7	✓	80 ms	$	✓
Regione 8	✓	15 ms	$	X

Tabella a matrice delle caratteristiche della regione

La tabella a matrice precedente mostra che la regione 6 è l'opzione migliore per questo scenario specifico perché la latenza è bassa rispetto ad altre regioni, il servizio è disponibile ed è la regione meno costosa.

Passaggi dell'implementazione

- Rivedi Regione AWS i prezzi: analizza i costi del carico di lavoro nella regione corrente. A partire dai costi più elevati per servizio e tipo di utilizzo, calcola i costi in altre regioni disponibili. Se il risparmio previsto supera il costo di spostamento del componente o del carico di lavoro, esegui la migrazione alla nuova regione.
- Rivedi i requisiti per implementazioni multi-regione: analizza i requisiti e gli obblighi aziendali (privacy dei dati, sicurezza o prestazioni) per scoprire se ci sono restrizioni che impediscono di utilizzare più regioni. Se non ci sono obblighi che limitano l'utilizzo di una sola regione, allora utilizza più regioni.
- Analizza il trasferimento dei dati necessario: nel selezionare le regioni, valuta i costi di trasferimento dei dati. Mantieni i dati vicino ai clienti e alle risorse. Scegli la soluzione meno costosa Regioni AWS dove scorrono i dati e dove il trasferimento dei dati è minimo. A seconda dei requisiti aziendali per il trasferimento dei dati, puoi utilizzare Amazon CloudFront e AWS Virtual

Private Networkper ridurre i costi di rete, migliorare le prestazioni e potenziare la sicurezza. AWS PrivateLinkAWS Direct Connect

Risorse

Documenti correlati:

- Accessing Reserved Instance recommendations
- EC2Prezzi Amazon
- Opzioni di acquisto delle istanze
- Tabella delle regioni

Video correlati:

- Save up to 90% and run production workloads on Spot

Esempi correlati:

- Overview of Data Transfer Costs for Common Architectures
- Cost Considerations for Global Deployments
- What to Consider when Selecting a Region for your Workloads
- Well-Architected Labs: limita l'utilizzo dei servizi per regione (Level 200)

COST07-BP03 Seleziona accordi di terze parti con condizioni convenienti

I contratti e i termini convenienti assicurano che i costi di questi servizi siano ridimensionati in base ai vantaggi che offrono. Seleziona i contratti e i prezzi che si ridimensionano quando forniscono ulteriori vantaggi alla tua organizzazione.

Livello di rischio associato se questa best practice non fosse adottata: medio

Guida all'implementazione

Sul mercato esistono diversi prodotti che possono aiutarti a gestire i costi negli ambienti cloud.
In termini di funzionalità possono presentare alcune differenze che dipendono dalle esigenze del cliente, ad esempio alcuni clienti sono più interessati alla governance o alla visibilità dei costi mentre

altri all'ottimizzazione di questi ultimi. Un fattore chiave per rendere più efficaci l'ottimizzazione e la governance dei costi è l'utilizzo dello strumento giusto con le funzionalità necessarie combinato al giusto modello di prezzo. Questi prodotti hanno modelli di prezzo diversi. Alcuni addebitano una determinata percentuale dell'importo fatturato mensilmente, mentre altri addebitano una percentuale dei risparmi realizzati. Idealmente, dovresti pagare solo ciò che hai effettivamente utilizzato.

Quando utilizzi soluzioni o servizi di terze parti nel cloud, è importante che le strutture dei prezzi siano allineate ai risultati desiderati. I prezzi devono essere scalati in base ai risultati e al valore che forniscono. Ad esempio, se utilizzi un software che contempla una percentuale del risparmio che fornisce, più risparmi (come risultato) e maggiore sarà l'importo addebitato. I contratti di licenza in cui paghi di più all'aumentare delle spese potrebbero non essere sempre nel tuo interesse ai fini dell'ottimizzazione dei costi. Tuttavia, se il fornitore offre vantaggi evidenti per tutte le voci incluse in fattura, questa tariffa scalare potrebbe essere giustificata.

Ad esempio, una soluzione che fornisce consigli per Amazon EC2 e addebita una percentuale dell'intera bolletta può diventare più costosa se utilizzi altri servizi che non offrono vantaggi. Un altro esempio è un servizio gestito che viene addebitato a una percentuale del costo delle risorse gestite. Una dimensione di istanza più grande potrebbe non richiedere necessariamente un maggiore impegno di gestione, ma potrebbe comportare un addebito superiore. Verifica che queste disposizioni tariffarie dei servizi includano un programma di ottimizzazione dei costi o funzionalità di servizio volte a migliorare l'efficienza.

I clienti potrebbero trovare i prodotti sul mercato più avanzati o più facili da usare. È necessario considerare il costo di questi prodotti e valutare i potenziali risultati di ottimizzazione dei costi a lungo termine.

Passaggi dell'implementazione

- Analizza contratti e termini stabiliti con terze parti: esamina i prezzi indicati nei contratti con terze parti. Esegui la modellazione per diversi livelli di utilizzo e considera i nuovi costi, come il nuovo utilizzo del servizio o aumenti dei servizi attuali a causa della crescita del carico di lavoro. Decidi se i costi aggiuntivi forniscono i vantaggi necessari alla tua azienda.

Risorse

Documenti correlati:

- Accessing Reserved Instance recommendations
- Opzioni di acquisto delle istanze

Video correlati:

- Save up to 90% and run production workloads on Spot

COST07-BP04 Implementa modelli di prezzo per tutti i componenti di questo carico di lavoro

Le risorse in esecuzione in modo permanente devono utilizzare la capacità riservata, ad esempio Savings Plans o istanze riservate. La capacità a breve termine è configurata per usare le istanze spot o il parco istanze spot. Le istanze on demand vengono utilizzate solo per carichi di lavoro a breve termine che non possono essere interrotti e che non durano abbastanza a lungo per la capacità riservata, tra il 25% e il 75% del periodo, a seconda del tipo di risorsa.

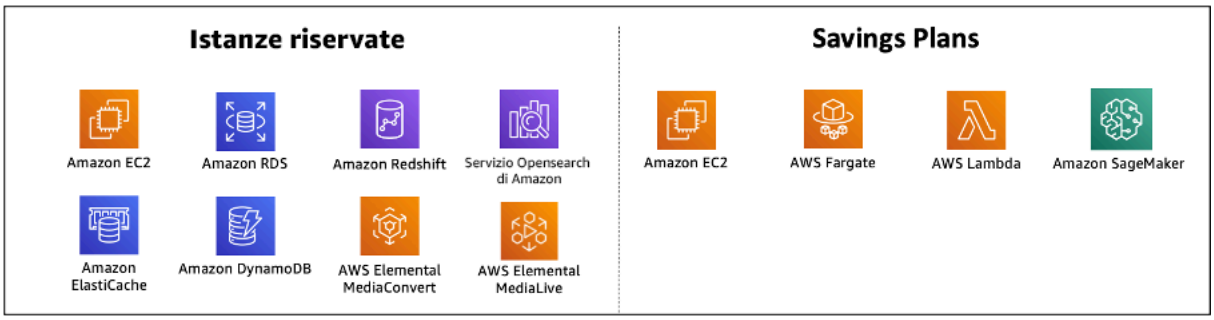
Livello di rischio associato se questa best practice non fosse adottata: basso

Guida all'implementazione

Per migliorare l'efficienza dei costi, AWS fornisce diversi consigli di impegno basati sull'utilizzo passato. Puoi utilizzare questi consigli per capire cosa puoi risparmiare e il livello di impegno richiesto. Puoi utilizzare questi servizi come On-Demand, Spot o impegnarti per un determinato periodo di tempo e ridurre i costi on-demand con Reserved Instances (RIs) e Savings Plans (). SPs Per ottimizzare il carico di lavoro, è necessario comprendere non solo i singoli componenti del carico di lavoro e AWS i diversi servizi, ma anche gli sconti vincolanti, le opzioni di acquisto e le istanze Spot per questi servizi.

Considera i requisiti dei componenti del tuo carico di lavoro e valuta i diversi modelli di prezzo per questi servizi. Definisci il requisito di disponibilità dei componenti. Determina l'eventuale presenza di più risorse indipendenti che eseguono la funzione nel carico di lavoro e quali sono i requisiti dello stesso nel corso del tempo. Confronta il costo delle risorse utilizzando il modello di prezzo on demand predefinito e altri modelli applicabili. Tieni conto di qualsiasi potenziale modifica nelle risorse o nei componenti del carico di lavoro.

Analizza, ad esempio, questa architettura di applicazione Web su AWS. Questo carico di lavoro di esempio è composto da più AWS servizi, come Amazon Route 53, Amazon AWS WAF, istanze Amazon CloudFront, EC2 istanze AmazonRDS, Load Balancers, storage Amazon S3 e Amazon Elastic File System (Amazon). EFS È necessario esaminare ciascuno di questi servizi e individuare le potenziali opportunità di risparmio sui costi con diversi modelli di prezzo. Alcuni di essi potrebbero essere idonei RIs oSPs, mentre altri potrebbero essere disponibili solo su richiesta. Come mostra l'immagine seguente, alcuni AWS servizi possono essere confermati utilizzando RIs oSPs.



AWS servizi impegnati utilizzando Reserved Instances e Savings Plans

Passaggi dell'implementazione

- Implementa modelli di prezzo: partendo dai risultati delle tue analisi, acquista Savings Plans, istanze riservate o implementa istanze spot. Se si tratta del primo acquisto con impegno, scegliete i cinque o dieci prodotti consigliati dall'elenco, quindi monitorate e analizzate i risultati nei prossimi mesi o due. AWS Cost Management Console ti guida attraverso il processo. Rivedi i consigli relativi all'istanza riservata (RI) o al modello Savings Plans sulla console, personalizza i consigli (tipo, pagamento e durata) e rivedi l'impegno orario (ad esempio, 20 USD all'ora), quindi aggiungilo al carrello. Gli sconti sono applicati automaticamente all'utilizzo idoneo. Acquista un importo ridotto di sconti a fronte di impegni a cicli regolari, ad esempio ogni 2 settimane o ogni mese. Implementa istanze spot per carichi di lavoro che possono essere interrotti o che sono stateless. Infine, seleziona le EC2 istanze Amazon on-demand e alloca le risorse per i requisiti rimanenti.
- Ciclo di revisione del carico di lavoro: implementa un ciclo di revisione per il carico di lavoro che analizzi in modo specifico la copertura del modello di prezzo. Quando il carico di lavoro ha la copertura necessaria, acquista ulteriori sconti a fronte di impegni parzialmente (ogni pochi mesi) o al variare dell'utilizzo dell'organizzazione.

Risorse

Documenti correlati:

- Understanding your Savings Plans recommendations
- Accessing Reserved Instance recommendations
- Modalità di acquisto delle istanze riservate
- Opzioni di acquisto delle istanze
- Spot Instances
- Modelli di prenotazione per altri servizi AWS

- Servizi supportati da Savings Plans

Video correlati:

- Save up to 90% and run production workloads on Spot

Esempi correlati:

- What should you consider before purchasing Savings Plans?
- How can I use Cost Explorer to analyze my spending and usage?

COST07-BP05 Eseguire l'analisi del modello di determinazione dei prezzi a livello di account di gestione

Verifica gli strumenti di gestione dei costi e di fatturazione e dai un'occhiata agli sconti suggeriti con impegni e prenotazioni per eseguire analisi regolari a livello di account di gestione.

Livello di rischio associato se questa best practice non fosse adottata: basso

Guida all'implementazione

L'esecuzione della modellazione dei costi a intervalli regolari garantisce l'implementazione di opportunità di ottimizzazione su più carichi di lavoro. Ad esempio, se più carichi di lavoro utilizzano istanze on demand a livello aggregato, il rischio di modifica è inferiore e l'implementazione di uno sconto a fronte di impegni permetterà di raggiungere un costo complessivo inferiore. Si consiglia di eseguire l'analisi a cicli regolari a cadenza quindicinale in un mese. In questo modo è possibile effettuare acquisti in piccoli incrementi, così che la copertura dei modelli di prezzo evolva di pari passo con i carichi di lavoro e i relativi componenti.

Usa lo strumento per i suggerimenti AWS Cost Explorer per scoprire opportunità di sconti a fronte di impegni nell'account di gestione. I suggerimenti a livello di account di gestione sono calcolati considerando l'utilizzo di tutti gli account nella tua organizzazione AWS che presenta istanze riservata (RI) o Savings Plans (SP) Vengono inoltre calcolati quando viene attivata la condivisione degli sconti per consigliare un impegno che massimizzi i risparmi su tutti gli account.

Sebbene l'acquisto a livello di account di gestione consenta di ottimizzare i risparmi in molti casi, in alcuni casi potresti prendere in considerazione l'acquisto SPs a livello di account collegato, ad esempio quando desideri che gli sconti si applichino prima all'utilizzo in quel particolare account collegato. I suggerimenti degli account membri sono calcolati a livello di singolo account per

massimizzare i risparmi per ogni account isolato. Se il tuo account ha vincoli o impegni sia per istanze riservate (RI) che per Savings Plans (SP), questi verranno applicati nel seguente ordine:

1. RI zonale
2. RI standard
3. RI convertibile
4. Piano di risparmio delle istanze
5. Piano di risparmio di calcolo

Se acquisti un SP a livello di account di gestione, i risparmi verranno applicati in base alla percentuale di sconto dalla più alta alla più bassa. SPsa livello di account di gestione, esamina tutti gli account collegati e applica i risparmi laddove lo sconto sarà il più alto. Se desideri limitare il luogo in cui vengono applicati i risparmi, puoi acquistare un Savings Plan a livello di account collegato e ogni volta che l'account esegue servizi di calcolo idonei, verrà applicato lo sconto. Quando l'account non esegue servizi di calcolo idonei, lo sconto verrà condiviso con gli altri account collegati con lo stesso account di gestione. La condivisione degli sconti è attivata per impostazione predefinita, ma può essere disattivata se necessario.

In una famiglia con fatturazione consolidata, i Savings Plans vengono applicati prima all'utilizzo dell'account del proprietario e, quindi, all'utilizzo degli altri account. Ciò si verifica solo se la condivisione è abilitata. I tuoi Savings Plans vengono applicati per primi alla percentuale di risparmio più alta. Se ci sono più utilizzi con percentuali di risparmio uguali, i Savings Plans vengono applicati al primo utilizzo con la tariffa dei Savings Plans più bassa. I Savings Plans continuano a essere validi fino all'esaurimento degli usi rimanenti o fino all'esaurimento del tuo impegno. L'eventuale utilizzo residuo viene addebitato in base alle tariffe on demand. Puoi aggiornare Savings Plans Recommendations in AWS Cost Management per generare nuovi Savings Plans Recommendations in qualsiasi momento.

Dopo aver analizzato la flessibilità delle istanze, puoi prendere una decisione in base ai suggerimenti ricevuti. Crea modelli di costo analizzando i costi a breve termine del carico di lavoro con potenziali diverse opzioni di risorse, analizzando i modelli di AWS prezzo e allineandoli ai requisiti aziendali per scoprire il costo totale di proprietà e le opportunità di ottimizzazione dei costi.

Passaggi dell'implementazione

Esegui un'analisi degli sconti a fronte di un impegno: con Cost Explorer nel tuo account, consulta i consigli su Savings Plans e istanze riservate. Verifica di aver compreso i suggerimenti dei Savings

Plans, fai una stima della tua spesa mensile e calcola il risparmio che puoi ottenere su tale intervallo di tempo. Esamina i consigli a livello di account di gestione, calcolati considerando l'utilizzo in tutti gli account membri della tua organizzazione AWS con abilitata la condivisione degli sconti Savings Plans o istanze riservate, per ottenere il massimo risparmio tra gli account. Per assicurarti di implementare le raccomandazioni corrette con gli sconti e i rischi richiesti, segui i Well-Architected Labs.

Risorse

Documenti correlati:

- Come funzionano i prezzi? AWS
- Opzioni di acquisto delle istanze
- Panoramica del Saving Plan
- Saving Plan recommendations
- Accessing Reserved Instance recommendations
- Understanding your Saving Plans recommendation
- In che modo i Savings Plans si applicano al tuo AWS utilizzo
- Savings Plans con fatture consolidate
- Attivazione della condivisione di sconti istanze riservate e Savings Plans

Video correlati:

- Save up to 90% and run production workloads on Spot

Esempi correlati:

- AWS Well-Architected Lab: modelli di prezzi (Level 200)
- AWS Well-Architected Labs: analisi dei modelli di prezzi (Level 200)
- Cosa devo considerare prima di acquistare un Savings Plan?
- How can I use rolling Savings Plans to reduce commitment risk?
- Quando usare le istanze Spot

COST8. In che modo pianifichi i costi per il trasferimento dei dati?

Assicurati di pianificare e monitorare i costi di trasferimento dei dati in modo da poter prendere decisioni sull'architettura per ridurre al minimo i costi. Una modifica piccola ma efficace dell'architettura può ridurre drasticamente i costi operativi nel tempo.

Best practice

- COST08-BP01 Eseguire la modellazione del trasferimento dei dati
- COST08-BP02 Seleziona i componenti per ottimizzare i costi di trasferimento dei dati
- COST08-BP03 Implementazione di servizi per ridurre i costi di trasferimento dei dati

COST08-BP01 Eseguire la modellazione del trasferimento dei dati

Raccogli i requisiti dell'organizzazione ed esegui la modellizzazione del trasferimento dei dati del carico di lavoro e di ciascuno dei suoi componenti. Questo identifica il punto di costo più basso per le sue attuali esigenze di trasferimento dei dati.

Livello di rischio associato se questa best practice non fosse adottata: elevato

Guida all'implementazione

Quando si progetta una soluzione nel cloud, i costi del trasferimento dei dati vengono in genere ignorati a causa dell'abitudine di progettare l'architettura utilizzando data center on-premises o per mancanza di conoscenze. I costi di trasferimento dei dati in AWS entrata sono determinati dalla fonte, dalla destinazione e dal volume del traffico. Tenere conto di questi costi durante la fase di progettazione può produrre risparmi. Capire dove avviene il trasferimento dei dati nell'ambito del carico di lavoro, il costo del trasferimento e i relativi vantaggi è molto importante per stimare con precisione il costo totale di proprietà (TCO). In questo modo puoi prendere una decisione consapevole quando si tratta di modificare o accettare una decisione relativa all'architettura. Ad esempio, potresti disporre di una configurazione con più zone di disponibilità dove replichi i dati tra le varie zone di disponibilità.

Puoi modellare i componenti dei servizi che trasferiscono i dati nel carico di lavoro e decidere che si tratta di un costo accettabile (simile a quello del calcolo e dell'archiviazione in entrambe le zone di disponibilità) per ottenere l'affidabilità e la resilienza richieste. Modella i costi in base a livelli differenti di utilizzo. L'utilizzo del carico di lavoro può cambiare nel corso del tempo e servizi differenti possono risultare più convenienti a livelli differenti.

Mentre modelli il trasferimento dei dati, pensa alla quantità di dati acquisiti e alla loro provenienza. Inoltre, considera la quantità di dati elaborati e la capacità di archiviazione o calcolo necessaria. Durante la modellazione, attieniti alle best practice relative alle reti in relazione all'architettura del carico di lavoro per ottimizzare i potenziali costi di trasferimento dei dati.

AWS Pricing Calculator Può aiutarti a visualizzare i costi stimati per AWS servizi specifici e il trasferimento di dati previsto. Se hai già un carico di lavoro in esecuzione (a scopo di test o in un ambiente di preproduzione), usa AWS Cost Explorero AWS Cost and Usage Report(CUR) per comprendere e modellare i costi di trasferimento dei dati. Configura un proof of concept (PoC) o testa il carico di lavoro ed esegui un test con un carico simulato realistico. Puoi modellare i costi in base alle diverse esigenze di carico di lavoro.

Passaggi dell'implementazione

- Identificazione dei requisiti: quali sono l'obiettivo principale e i requisiti aziendali per il trasferimento pianificato dei dati tra origine e destinazione? Qual è il risultato aziendale previsto finale? Acquisisci i requisiti aziendali e definisci il risultato previsto.
- Identifica origine e destinazione: quali sono l'origine e la destinazione dei dati per il trasferimento dei dati, ad esempio all'interno Regioni AWS, verso AWS i servizi o verso Internet?
  - Trasferimento dei dati all'interno di un Regione AWS
  - Trasferimento di dati tra Regioni AWS
  - Trasferimento di dati verso Internet
- Identificazione delle classificazioni dei dati: qual è la classificazione dei dati per il trasferimento di dati in questione? Di che tipo di dati si tratta? Quali sono le dimensioni dei dati? Con quale frequenza devono essere trasferiti i dati? I dati sono sensibili?
- Identifica AWS i servizi o gli strumenti da utilizzare: quali AWS servizi vengono utilizzati per questo trasferimento di dati? È possibile utilizzare un servizio già allocato a un altro carico di lavoro?
- Calcolo dei costi di trasferimento dei dati: utilizza i prezzi AWS, nonché il modello di trasferimento dati creato in precedenza, per calcolare i costi di trasferimento dei dati per il carico di lavoro. Calcola i costi di trasferimento dei dati a diversi livelli di utilizzo, ipotizzando incrementi e riduzioni dell'utilizzo del carico di lavoro. Nei casi in cui sono disponibili più opzioni per l'architettura del carico di lavoro valuta i costi di ogni opzione per il confronto.
- Collegamento dei costi ai risultati: per ogni costo di trasferimento dei dati sostenuto, specifica il risultato ottenuto per il carico di lavoro. Se si tratta di un trasferimento tra componenti potrebbe trattarsi di una necessità di disaccoppiamento, se si tratta di un trasferimento tra zone di disponibilità potrebbe trattarsi di una necessità di ridondanza.

- Creazione della modellazione per il trasferimento dei dati: una volta raccolte tutte le informazioni, crea una base concettuale di modellazione del trasferimento dei dati per più casi d'uso e diversi carichi di lavoro.

Risorse

Documenti correlati:

- AWS soluzioni di memorizzazione nella cache
- AWS Prezzi
- EC2Prezzi Amazon
- VPCPrezzi Amazon
- Understanding data transfer charges

Video correlati:

- Monitoring and Optimizing Your Data Transfer Costs
- S3 Transfer Acceleration

Esempi correlati:

- Overview of Data Transfer Costs for Common Architectures
- AWS Linee guida prescrittive per il networking

COST08-BP02 Seleziona i componenti per ottimizzare i costi di trasferimento dei dati

Tutti i componenti sono selezionati e l'architettura è progettata per ridurre i costi di trasferimento dei dati. Ciò include l'utilizzo di componenti come wide-area-network (WAN) ottimizzazione e configurazioni Multi-Availability Zone (AZ)

Livello di rischio associato se questa best practice non fosse adottata: medio

Guida all'implementazione

Una progettazione basata sul trasferimento dei dati riduce i costi del trasferimento stesso. Potrebbe implicare l'uso di reti di distribuzione di contenuti per posizionare i dati vicino agli utenti, oppure l'uso di collegamenti di rete dedicati dalle tue sedi ad AWS. È inoltre possibile utilizzare

l'WANottimizzazione e l'ottimizzazione delle applicazioni per ridurre la quantità di dati trasferiti tra i componenti.

Quando si trasferiscono dati verso o all'interno di Cloud AWS, è essenziale conoscere la destinazione in base ai vari casi d'uso, la natura dei dati e le risorse di rete disponibili per selezionare AWS i servizi giusti per ottimizzare il trasferimento dei dati. AWS offre una gamma di servizi di trasferimento dati personalizzati per diverse esigenze di migrazione dei dati. Seleziona le opzioni di archiviazione di dati e trasferimento di dati opportune in base alle esigenze aziendali all'interno della tua organizzazione.

Quando pianifichi o rivedi l'architettura di un carico di lavoro, considera quanto segue:

- Usa VPC gli endpoint within AWS: gli VPC endpoint consentono connessioni private tra i tuoi servizi VPC e quelli supportati AWS . Ciò consente di evitare l'utilizzo della rete Internet pubblica, che può comportare costi di trasferimento dei dati.
- Usa un NAT gateway: utilizza un NATgateway in modo che le istanze in una sottorete privata possano connettersi a Internet o ai servizi esterni al tuo. VPC Verifica se le risorse del NAT gateway che inviano la maggior parte del traffico si trovano nella stessa zona di disponibilità del NAT gateway. In caso contrario, crea nuovi NAT gateway nella stessa zona di disponibilità della risorsa per ridurre i costi di trasferimento dati Cross-AZ.
- Use AWS Direct Connect AWS Direct Connect ignora la rete Internet pubblica e stabilisce una connessione diretta e privata tra la rete locale e. AWS Ciò può essere più conveniente e coerente rispetto al trasferimento di grandi volumi di dati su Internet.
- Evita di trasferire dati attraverso i confini regionali: i trasferimenti di dati tra Regioni AWS (da una regione all'altra) in genere comportano costi. Seguire questo approccio basato sul trasferimento tra regioni dovrebbe essere una decisione molto ponderata. Per ulteriori informazioni, consulta Scenari multi-regione.
- Monitora il trasferimento dei dati: utilizza Amazon CloudWatch e i log di VPC flusso per acquisire dettagli sul trasferimento dei dati e sull'utilizzo della rete. Analizza le informazioni sul traffico di rete acquisite nel tuo VPCs computer, ad esempio l'indirizzo IP o l'intervallo da e verso le interfacce di rete.
- Analizza l'utilizzo della rete: utilizza strumenti di misurazione e reportistica come AWS Cost Explorer i CUDOS dashboard o CloudWatch per comprendere i costi di trasferimento dei dati del carico di lavoro.

Passaggi dell'implementazione

- Seleziona i componenti per il trasferimento dei dati: utilizzando la modellazione per il trasferimento dei dati illustrata in COST08-BP01 Eseguire la modellazione del trasferimento dei dati, concentrati su dove si trovano i costi di trasferimento dei dati più elevati o dove sarebbero se l'utilizzo del carico di lavoro cambiasse. Individua architetture alternative o componenti aggiuntivi che eliminano o riducono la necessità di trasferimento dei dati o ne riducono i costi.

Risorse

Best practice correlate:

- COST08-BP01 Eseguire la modellazione del trasferimento dei dati
- COST08-BP03 Implementazione di servizi per ridurre i costi di trasferimento dei dati

Documenti correlati:

- Migrazione dei dati nel cloud
- AWS caching solutions
- Distribuisci contenuti più velocemente con Amazon CloudFront

Esempi correlati:

- Overview of Data Transfer Costs for Common Architectures
- AWS Suggerimenti per l'ottimizzazione della rete
- Ottimizza le prestazioni e riduci i costi per l'analisi di rete con VPC Flow Logs in formato Apache Parquet

COST08-BP03 Implementazione di servizi per ridurre i costi di trasferimento dei dati

Implementa i servizi per ridurre il costo di trasferimento dei dati. Ad esempio, utilizzate le edge location o le reti di distribuzione dei contenuti (CDN) per distribuire contenuti agli utenti finali, create livelli di caching davanti ai server o ai database delle applicazioni e utilizzate connessioni di rete dedicate anziché VPN per la connettività al cloud.

Livello di rischio associato se questa best practice non fosse adottata: medio

Guida all'implementazione

Esistono vari AWS servizi che possono aiutarvi a ottimizzare l'utilizzo del trasferimento di dati in rete. A seconda dei componenti del carico di lavoro, del tipo e dell'architettura cloud, questi servizi possono aiutarti nella compressione, nella memorizzazione nella cache, nella condivisione e distribuzione del traffico sul cloud.

- Amazon CloudFront è una rete globale di distribuzione di contenuti che fornisce dati con bassa latenza e velocità di trasferimento elevate. Memorizza nella cache i dati nelle posizioni edge di tutto il mondo, riducendo così il carico sulle tue risorse. Utilizzandola CloudFront, puoi ridurre lo sforzo amministrativo necessario per distribuire contenuti a un gran numero di utenti in tutto il mondo con una latenza minima. Il pacchetto di risparmio sulla sicurezza può aiutarti a risparmiare fino al 30% sull' CloudFront utilizzo se prevedi di aumentare l'utilizzo nel tempo.
- AWS Direct Connect ti consente di creare una connessione di rete dedicata ad AWS. In questo modo puoi ridurre i costi di rete, aumentare la larghezza di banda e offrire un'esperienza di rete più costante rispetto alle connessioni Internet.
- AWS VPN consente di stabilire una connessione sicura e privata tra la rete privata e la rete globale AWS . È ideale per piccoli uffici o partner aziendali perché offre una connettività semplificata ed è un servizio completamente gestito ed elastico.
- VPCGli endpoint consentono la connettività tra AWS i servizi tramite reti private e possono essere utilizzati per ridurre i costi del trasferimento pubblico dei dati e dei NATgateway. Gli VPCendpoint gateway non hanno costi orari e supportano Amazon S3 e Amazon DynamoDB. Gli VPCendpoint di interfaccia sono forniti da AWS PrivateLinke hanno una tariffa oraria e un costo di utilizzo per GB.
- NATI gateway offrono scalabilità e gestione integrate per ridurre i costi rispetto a un'istanza autonoma. NAT Posiziona i NAT gateway nelle stesse zone di disponibilità delle istanze ad alto traffico e valuta la possibilità di utilizzare gli VPC endpoint per le istanze che devono accedere ad Amazon DynamoDB o Amazon S3 per ridurre i costi di trasferimento ed elaborazione dei dati.
- Utilizza AWS Snow Familydispositivi dotati di risorse di elaborazione per raccogliere ed elaborare dati all'edge. AWS Snow Family i dispositivi (Snowcone, Snowball e Snowmobile) consentono di trasferire petabyte di dati in modo conveniente e offline. Cloud AWS

Passaggi dell'implementazione

- Implementazione dei servizi: seleziona i servizi di AWS rete applicabili in base al tipo di carico di lavoro del servizio utilizzando la modellazione del trasferimento dei dati e la revisione dei log

di flusso. VPC Scopri dove si trovano i costi maggiori e i flussi con volumi più elevati. Esamina i AWS servizi e valuta se esiste un servizio che riduce o rimuove il trasferimento, in particolare il networking e la distribuzione dei contenuti. Individua anche servizi di caching in cui si verifica un accesso ripetuto ai dati o in cui sono presenti grandi quantità di dati.

Risorse

Documenti correlati:

- AWS Direct Connect
- AWS Esplora i nostri prodotti
- AWS caching solutions
- Amazon CloudFront
- AWS Snow Family
- Pacchetto Amazon CloudFront Security Savings

Video correlati:

- Monitoring and Optimizing Your Data Transfer Costs
- AWS Serie di ottimizzazione dei costi: CloudFront
- Come posso ridurre i costi di trasferimento dati per il mio NAT gateway?

Esempi correlati:

- How-to chargeback shared services: An AWS Transit Gateway example
- Comprendi in modo approfondito i dettagli del trasferimento AWS dei dati dal rapporto sui costi e sull'utilizzo utilizzando la query Athena e QuickSight
- Overview of Data Transfer Costs for Common Architectures
- Using AWS Cost Explorer to analyze data transfer costs
- Ottimizzazione dei costi delle AWS architetture utilizzando le funzionalità di Amazon CloudFront
- Come posso ridurre i costi di trasferimento dati per il mio gateway? NAT

Gestione delle risorse di domanda e offerta

Domanda

- COST9. Come gestisci la domanda e fornisci le risorse?

COST9. Come gestisci la domanda e fornisci le risorse?

Per avere un carico di lavoro con costo e prestazioni bilanciate, verifica che venga utilizzato tutto ciò per cui paghi ed evita le istanze molto sottoutilizzate. Una metrica di utilizzo distorta in entrambe le direzioni ha un impatto negativo sull'organizzazione, in termini di costi operativi (riduzione delle prestazioni dovuta a un utilizzo eccessivo) o di AWS spese sprecate (a causa dell'over-provisioning).

Best practice

- COST09-BP01 Eseguire un'analisi sulla domanda del carico di lavoro
- COST09-BP02 Implementare un buffer o un acceleratore per gestire la domanda
- COST09-BP03 Fornisci risorse in modo dinamico

COST09-BP01 Eseguire un'analisi sulla domanda del carico di lavoro

Analizza la domanda del carico di lavoro nel tempo. Verifica che l'analisi copra l'andamento stagionale e rappresenti accuratamente le condizioni operative per l'intera durata del carico di lavoro. L'attività di analisi deve riflettere i potenziali benefici, ad esempio che il tempo speso sia proporzionale al costo del carico di lavoro.

Livello di rischio associato se questa best practice non fosse adottata: elevato

Guida all'implementazione

L'analisi della domanda di carichi di lavoro per il cloud computing implica la comprensione dei modelli e delle caratteristiche delle attività di elaborazione avviate nell'ambiente cloud. Questa analisi aiuta gli utenti a ottimizzare l'allocazione delle risorse, gestire i costi e verificare che le prestazioni soddisfino i livelli richiesti.

Scopri i requisiti del carico di lavoro. I requisiti dell'organizzazione devono indicare i tempi di risposta del carico di lavoro per le richieste. Il tempo di risposta può essere utilizzato per determinare se la domanda è gestita o se l'offerta di risorse cambierà per soddisfare la domanda.

L'analisi deve includere la prevedibilità e la ripetibilità della domanda, la velocità di variazione della domanda e la quantità di variazione della domanda. Esegui l'analisi per un periodo sufficientemente lungo da incorporare eventuali variazioni stagionali, ad esempio i picchi di elaborazione o le festività. end-of-month

Lo sforzo di analisi dovrebbe riflettere i potenziali vantaggi dell'implementazione della scalabilità. Osserva il costo totale previsto del componente ed eventuali aumenti o riduzioni di utilizzo e costi durante il ciclo di vita del carico di lavoro.

Di seguito sono riportati alcuni aspetti chiave da prendere in considerazione quando si esegue l'analisi della domanda del carico di lavoro per il cloud computing:

1. Utilizzo delle risorse e metriche delle prestazioni: analizza come le AWS risorse vengono utilizzate nel tempo. Determina i modelli di utilizzo di picco e non di picco per ottimizzare l'allocazione delle risorse e le strategie di scalabilità. Monitora le metriche delle prestazioni come tempi di risposta, latenza, throughput e tassi di errore. Queste metriche aiutano a valutare lo stato e l'efficienza complessive dell'infrastruttura cloud.
2. Comportamento in termini di dimensionamento di utenti e applicazioni: analizza il comportamento degli utenti e il relativo impatto sulla domanda del carico di lavoro. L'esame dei modelli di traffico degli utenti aiuta a migliorare la fornitura di contenuti e la reattività delle applicazioni. Analizza la modalità di dimensionamento dei carichi di lavoro in base all'aumento della domanda. Determina se i parametri di dimensionamento automatico sono configurati correttamente ed efficacemente per gestire le fluttuazioni del carico.
3. Tipi di carico di lavoro: identifica i diversi tipi di carichi di lavoro in esecuzione nel cloud, come l'elaborazione in batch, l'elaborazione dei dati in tempo reale, le applicazioni Web, i database o i processi di machine learning. Ogni tipo di carico di lavoro può avere requisiti di risorse e profili di prestazioni diversi.
4. Accordi sui livelli di servizio (SLAs): confronta le prestazioni effettive con SLAs per garantire la conformità e identifica le aree che necessitano di miglioramento.

Puoi usare Amazon CloudWatch per raccogliere e tracciare metriche, monitorare i file di registro, impostare allarmi e reagire automaticamente ai cambiamenti nelle tue AWS risorse. Puoi anche utilizzare Amazon CloudWatch per ottenere visibilità a livello di sistema sull'utilizzo delle risorse, sulle prestazioni delle applicazioni e sullo stato operativo.

Con AWS Trusted Advisor, puoi allocare le tue risorse seguendo le best practice per migliorare le prestazioni e l'affidabilità del sistema, aumentare la sicurezza e trovare opportunità di risparmio di

denaro. Puoi anche disattivare le istanze non di produzione e utilizzare Amazon CloudWatch e Auto Scaling per far fronte agli aumenti o alle riduzioni della domanda.

Infine, puoi utilizzare AWS Cost Explorero Amazon QuickSight con il file AWS Cost and Usage Report (CUR) o i log delle applicazioni per eseguire un'analisi avanzata della domanda del carico di lavoro.

Nel complesso, un'analisi completa della domanda dei carichi di lavoro consente alle organizzazioni di prendere decisioni informate sul provisioning, il dimensionamento e l'ottimizzazione delle risorse, con conseguente miglioramento delle prestazioni, dell'efficienza dei costi e della soddisfazione degli utenti.

Passaggi dell'implementazione

- Analizza i carichi di lavoro esistenti: analizza i dati provenienti dal carico di lavoro esistente, dalle versioni precedenti del carico di lavoro o dai modelli di utilizzo previsti. Usa Amazon CloudWatch, i file di log e i dati di monitoraggio per ottenere informazioni dettagliate su come è stato utilizzato il carico di lavoro. Analizza un ciclo completo del carico di lavoro e raccogli dati per eventuali cambiamenti stagionali, ad esempio eventi end-of-month. end-of-year L'attività che emerge dall'analisi deve riflettere le caratteristiche del carico di lavoro. L'impegno maggiore dovrebbe riguardare i carichi di lavoro di alto valore che presentano le maggiori variazioni della domanda. Il minimo impegno dovrebbe riguardare carichi di lavoro di basso valore che hanno variazioni minime nella domanda.
- Prevedi le influenze esterne: incontra i membri del team di tutta l'organizzazione che possono influenzare o modificare la domanda del carico di lavoro. I team più comuni sono le vendite, il marketing o il business development. Collabora con loro per conoscere i cicli secondo cui operano e se ci sono eventi che potrebbero modificare la domanda del carico di lavoro. Prevedi la richiesta del carico di lavoro con questi dati.

Risorse

Documenti correlati:

- Amazon CloudWatch
- AWS Trusted Advisor
- AWS X-Ray
- AWS Auto Scaling

- AWS Instance Scheduler
- Guida introduttiva ad Amazon SQS
- AWS Cost Explorer
- Amazon QuickSight

Video correlati:

Esempi correlati:

- Monitor, Track and Analyze for cost optimization
- Ricerca e analisi dei log in CloudWatch

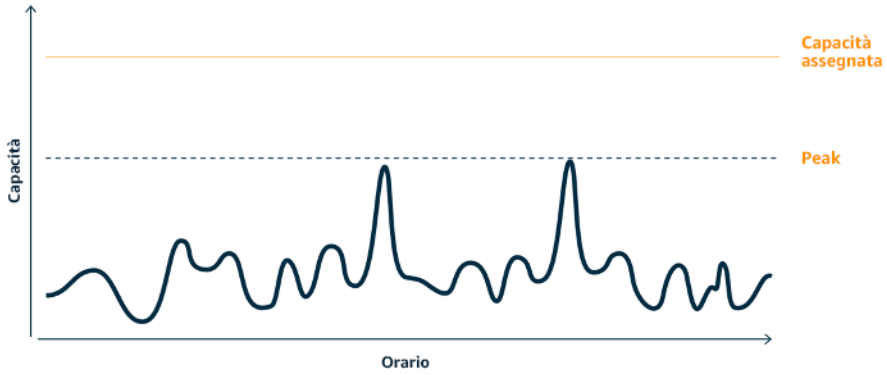
COST09-BP02 Implementare un buffer o un acceleratore per gestire la domanda

Il buffering e la limitazione (della larghezza di banda della rete) modificano la domanda sul carico di lavoro, attenuando eventuali picchi. Implementa la limitazione (della larghezza di banda della rete) quando i client eseguono nuovi tentativi. Implementa il buffering per archiviare la richiesta e rinviare l'elaborazione a un secondo momento. Verifica che le esecuzioni di limitazione (della larghezza di banda della rete) e buffering siano progettate in modo che i client ricevano una risposta nel tempo richiesto.

Livello di rischio associato se questa best practice non fosse adottata: medio

Guida all'implementazione

L'implementazione di un buffer o di una limitazione (della larghezza di banda della rete) è fondamentale nel cloud computing per gestire la domanda e ridurre la capacità allocata richiesta per il carico di lavoro. Per ottenere prestazioni ottimali, è essenziale valutare la domanda totale, compresi i picchi, la velocità con cui variano le richieste e il tempo di risposta necessario. Quando i client hanno la possibilità di inviare nuovamente le proprie richieste, conviene applicare la limitazione (della larghezza di banda della rete). Al contrario, per i client che non dispongono della funzionalità di esecuzione di nuovi tentativi, l'approccio ideale è implementare una soluzione buffer. Tali buffer semplificano l'afflusso di richieste e ottimizzano l'interazione delle applicazioni con diverse velocità operative.



Curva di domanda con due picchi distinti che richiedono un'elevata capacità allocata

Supponiamo che un carico di lavoro sia caratterizzato dalla curva della domanda illustrata nella figura precedente. Questo carico di lavoro presenta due picchi e per gestire tali picchi viene eseguito il provisioning della capacità di risorse mostrata dalla linea arancione. Le risorse e l'energia utilizzate per questo carico di lavoro non sono indicate nell'area sotto la curva della domanda, ma nell'area sotto la linea della capacità allocata, poiché per gestire questi due picchi è necessario eseguire il provisioning di tale capacità. Diminuire la curva della domanda del carico di lavoro può aiutarti a ridurre la capacità allocata di un carico di lavoro, oltre al suo impatto sull'ambiente. Per attenuare il picco, valuta la possibilità di implementare una soluzione basata sulla limitazione (della larghezza di banda della rete) o sul buffering.

Per comprendere meglio queste buffering e limitazione (della larghezza di banda della rete), proviamo ad analizzarle.

Limitazione (della larghezza di banda della rete): se l'origine della richiesta dispone di funzionalità di ripetizione dei tentativi, è possibile implementare la limitazione (della larghezza di banda della rete). La limitazione (della larghezza di banda della rete) indica all'origine che, se non è in grado di soddisfare la richiesta all'ora corrente, dovrebbe riprovare più tardi. L'origine attende un periodo di tempo, quindi riprova a eseguire la richiesta. L'implementazione della limitazione (della larghezza di banda della rete) ha il vantaggio di limitare la quantità massima di risorse e i costi del carico di lavoro. Nel AWS, puoi utilizzare Amazon API Gateway per implementare il throttling.

Basato sul buffer: un approccio basato sul buffer si appoggia a produttori (componenti che inviano messaggi alla coda), consumatori (componenti che ricevono messaggi dalla coda) e una coda (che contiene messaggi) per l'archiviazione dei messaggi. I messaggi vengono letti ed elaborati dai consumatori e ciò consente ai messaggi di essere eseguiti alla velocità che soddisfa i requisiti aziendali del consumatore stesso. Utilizzando una metodologia basata sul buffering, i messaggi dei

produttori sono ospitati in code o flussi, dove i produttori possono accedervi a un ritmo in linea con le rispettive esigenze operative.

Nel AWS, puoi scegliere tra più servizi per implementare un approccio di buffering. Amazon Simple Queue Service (AmazonSQS) è un servizio gestito che fornisce code che consentono a un singolo consumatore di leggere singoli messaggi. Amazon Kinesis offre un flusso che consente a più consumatori di leggere gli stessi messaggi.

Il buffering e la limitazione (della larghezza di banda della rete) possono attenuare eventuali picchi modificando la domanda sul carico di lavoro. Usa la limitazione (della larghezza di banda della rete) quando i client riprovano le azioni e usa il buffering per bloccare la richiesta ed elaborarla in un secondo momento. Durante l'utilizzo dell'approccio basato sul buffering, assicurati di progettare il carico di lavoro per soddisfare la richiesta nel tempo richiesto e verifica di essere in grado di gestire le richieste duplicate. Analizza la domanda complessiva, la velocità di modifica e il tempo di risposta richiesto per determinare le dimensioni della limitazione (della larghezza di banda della rete) o del buffer richiesto.

Passaggi dell'implementazione

- Analizza i requisiti del client: analizza le richieste del client per determinare se sono in grado di eseguire nuovi tentativi. Per i client che non possono eseguire nuovi tentativi, è necessario implementare i buffer. Analizza la domanda complessiva, la velocità di modifica e il tempo di risposta richiesto per determinare le dimensioni della limitazione (della larghezza di banda della rete) o del buffer richiesto.
- Implementa un buffer o una limitazione (della larghezza di banda della rete): implementa un buffer o una limitazione (della larghezza di banda della rete) nel carico di lavoro. Una coda come Amazon Simple Queue Service SQS (Amazon) può fornire un buffer ai componenti del carico di lavoro. Amazon API Gateway può fornire la limitazione dei componenti del carico di lavoro.

Risorse

Best practice correlate:

- SUS02-BP06 Implementa il buffering o il throttling per appiattire la curva di domanda
- REL05-BP02 Richieste Throttle

Documenti correlati:

- AWS Auto Scaling
- AWS Instance Scheduler
- Amazon API Gateway
- Amazon Simple Queue Service
- Guida introduttiva ad Amazon SQS
- Amazon Kinesis

Video correlati:

- Choosing the Right Messaging Service for Your Distributed App

Esempi correlati:

- Gestione e monitoraggio della API limitazione dei carichi di lavoro
- Limitazione su larga scala di un sistema multi-tenant su più livelli utilizzando Gateway REST API API
- Abilitazione del tiering e del throttling in una soluzione Amazon SaaS multi-tenant utilizzando Amazon EKS Gateway API
- Application integration Using Queues and Messages

COST09-BP03 Fornisci risorse in modo dinamico

Le risorse sono allocate in modo pianificato. La pianificazione può essere basata sulla domanda, ad esempio tramite il dimensionamento automatico, oppure sul tempo, quando la domanda è prevedibile e le risorse sono fornite in base al tempo. Questi metodi comportano la minore quantità possibile di provisioning in eccesso o in difetto.

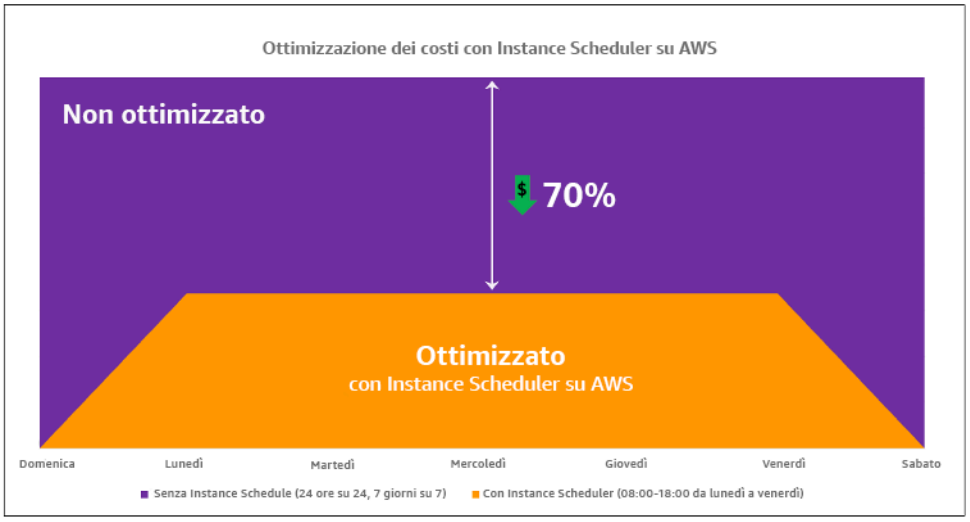
Livello di rischio associato se questa best practice non fosse adottata: basso

Guida all'implementazione

I AWS clienti possono aumentare le risorse disponibili per le proprie applicazioni e fornire risorse per soddisfare la domanda. Una di queste opzioni consiste nell'utilizzare AWS Instance Scheduler, che automatizza l'avvio e l'arresto delle istanze Amazon Elastic Compute Cloud (AmazonEC2) e Amazon Relational Database Service (Amazon). RDS L'altra opzione è l'utilizzo AWS Auto Scaling, che consente di scalare automaticamente le risorse di elaborazione in base alla domanda

dell'applicazione o del servizio. Fornire risorse in base alla domanda ti consentirà di pagare solo per le risorse che usi, di ridurre i costi lanciando le risorse quando sono necessarie e di interromperle quando non servono più.

AWS Instance Scheduler ti consente di configurare l'arresto e l'avvio delle RDS istanze Amazon EC2 e Amazon in orari definiti in modo da soddisfare la domanda delle stesse risorse entro uno schema temporale coerente, ad esempio ogni giorno gli utenti accedono alle EC2 istanze Amazon alle otto del mattino e non ne hanno bisogno dopo le sei di sera. Questa soluzione aiuta a ridurre i costi operativi fermando le risorse non utilizzate e avviandole quando sono necessarie.



Ottimizzazione dei costi con AWS Instance Scheduler.

Puoi anche configurare facilmente le pianificazioni per le tue EC2 istanze Amazon tra i tuoi account e le tue regioni con una semplice interfaccia utente (UI) utilizzando AWS Systems Manager Quick Setup. Puoi pianificare RDS istanze Amazon EC2 o Amazon con AWS Instance Scheduler e interrompere e avviare le istanze esistenti. Tuttavia, non puoi fermare e avviare istanze che fanno parte del tuo gruppo Auto Scaling ASG () o che gestiscono servizi come Amazon Redshift o Amazon Service. OpenSearch I gruppi Auto Scaling presentano una propria pianificazione in merito alle istanze del gruppo e queste istanze vengono create.

AWS Auto Scaling ti aiuta a regolare la capacità per mantenere prestazioni stabili e prevedibili al minor costo possibile per soddisfare le mutevoli esigenze. È un servizio completamente gestito e gratuito per scalare la capacità della tua applicazione che si integra con EC2 istanze Amazon e flotte Spot, Amazon, Amazon ECS DynamoDB e Amazon Aurora. Auto Scaling fornisce il rilevamento automatico delle risorse per aiutare a trovare risorse nel carico di lavoro che possono essere configurate, dispone di strategie di dimensionamento integrate per ottimizzare le prestazioni, i costi o

un equilibrio tra i due e fornisce il dimensionamento predittivo per aiutare a risolvere i picchi ricorrenti con regolarità.

Sono disponibili diverse opzioni di dimensionamento per scalare il tuo gruppo Auto Scaling:

- Mantenimento dei livelli di istanza correnti in qualsiasi momento
- Dimensionamento manuale
- Dimensionamento in base a una pianificazione
- Dimensionamento on demand
- Utilizzo del dimensionamento predittivo

Le policy di Auto Scaling sono diverse e possono essere classificate come policy di dimensionamento dinamico e pianificato. Le policy dinamiche fanno riferimento al dimensionamento manuale o dinamico, programmato o predittivo. È possibile utilizzare le policy di dimensionamento per il dimensionamento dinamico, pianificato e predittivo. Puoi anche utilizzare i parametri e gli allarmi di Amazon CloudWatch per attivare eventi di scalabilità per il tuo carico di lavoro. Noi ti suggeriamo di utilizzare i modelli di avvio, che consentono di accedere alle funzionalità e ai miglioramenti più recenti. In caso di utilizzo di configurazioni di avvio, non tutte le funzionalità di Auto Scaling sono disponibili. Ad esempio, non è possibile creare un gruppo Auto Scaling che avvii istanze spot e on demand oppure che specifichi più tipi di istanza. Per configurare queste caratteristiche, sarà necessario utilizzare un modello di avvio. Quando utilizzi i modelli di avvio, ti consigliamo di modificare ciascuno di essi. Con il controllo delle versioni dei modelli di avvio, è possibile creare un sottoinsieme del set completo di parametri. Quindi, è possibile riutilizzarlo per creare altre versioni dello stesso modello di avvio.

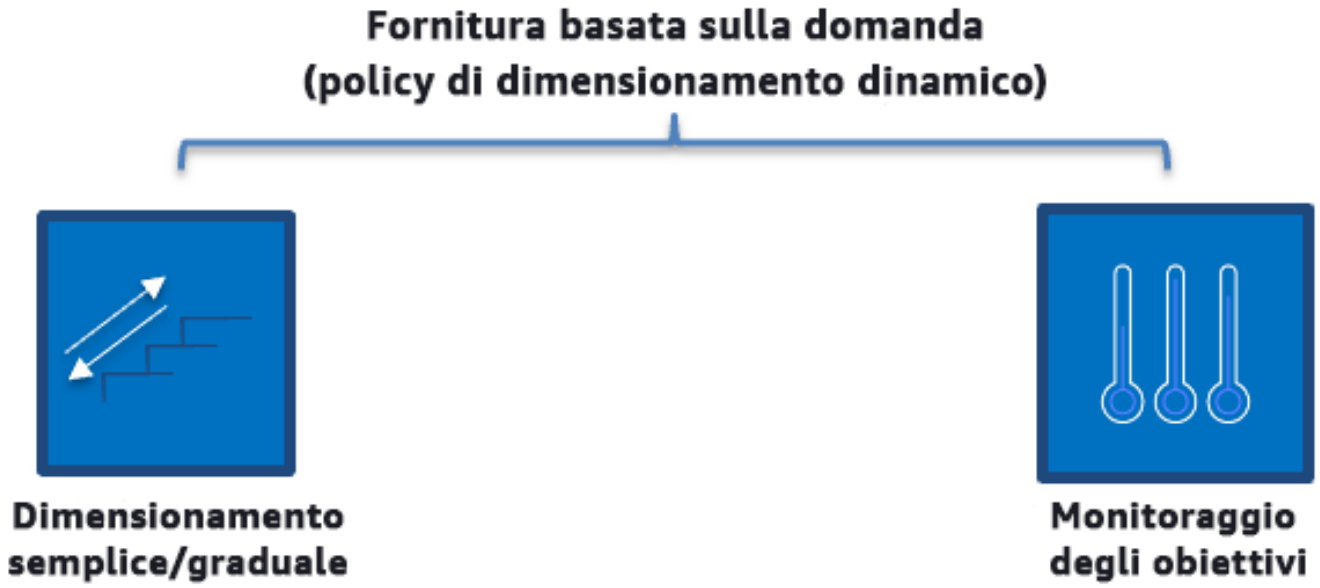
Puoi utilizzare AWS Auto Scaling o incorporare la scalabilità nel tuo codice con o.AWS APIs SDKs Ciò riduce i costi complessivi del carico di lavoro rimuovendo i costi operativi dall'apportare manualmente modifiche al tuo ambiente; le modifiche possono essere apportate molto più rapidamente. In questo modo, inoltre, il carico di lavoro viene adattato alla domanda in qualsiasi momento. Per seguire questa best practice e fornire risorse in modo dinamico alla tua organizzazione, devi comprendere la scalabilità orizzontale e verticale delle applicazioni in esecuzione sulle Cloud AWS istanze Amazon e la natura delle applicazioni in esecuzione su istanze AmazonEC2. È meglio che il team di Cloud Financial Management collabori con i team tecnici per seguire questa best practice.

Elastic Load Balancing (bilanciamento del carico elastico) consente di scalare le risorse distribuendo la domanda su più risorse. Con l'utilizzo ASG di Elastic Load Balancing, puoi gestire le richieste in

entrata instradando il traffico in modo ottimale in modo che nessuna istanza sia sovraccarica in un gruppo di Auto Scaling. Le richieste vengono distribuite tra tutti gli obiettivi di un gruppo target in modalità Round Robin, senza tenere conto della capacità o dell'utilizzo.

Le metriche tipiche possono essere parametri standard di AmazonEC2, come l'CPUutilizzo, il throughput di rete e la latenza di richiesta e risposta osservata da Elastic Load Balancing. Quando possibile, è consigliabile utilizzare un parametro indicativo dell'esperienza del cliente, in genere si tratta di un parametro personalizzato che potrebbe avere origine dal codice dell'applicazione all'interno del carico di lavoro. Per capire come soddisfare la domanda in modo dinamico in questo documento, Auto Scaling verrà suddiviso in due categorie (modello di fornitura basata sulla domanda e modello di fornitura basata sul tempo) e verrà approfondito ciascun modello.

Fornitura basata sulla domanda: sfrutta l'elasticità del cloud per fornire risorse in grado di soddisfare la domanda in continua evoluzione facendo riferimento allo stato della domanda quasi in tempo reale. Per funzionalità di fornitura, utilizzo APIs o servizio basate sulla domanda per variare in modo programmatico la quantità di risorse cloud presenti nell'architettura. Ciò ti consente di scalare i componenti nella tua architettura e aumentare il numero di risorse durante i picchi di domanda per mantenere le prestazioni, nonché diminuire la capacità quando la domanda cala in modo da ridurre i costi.



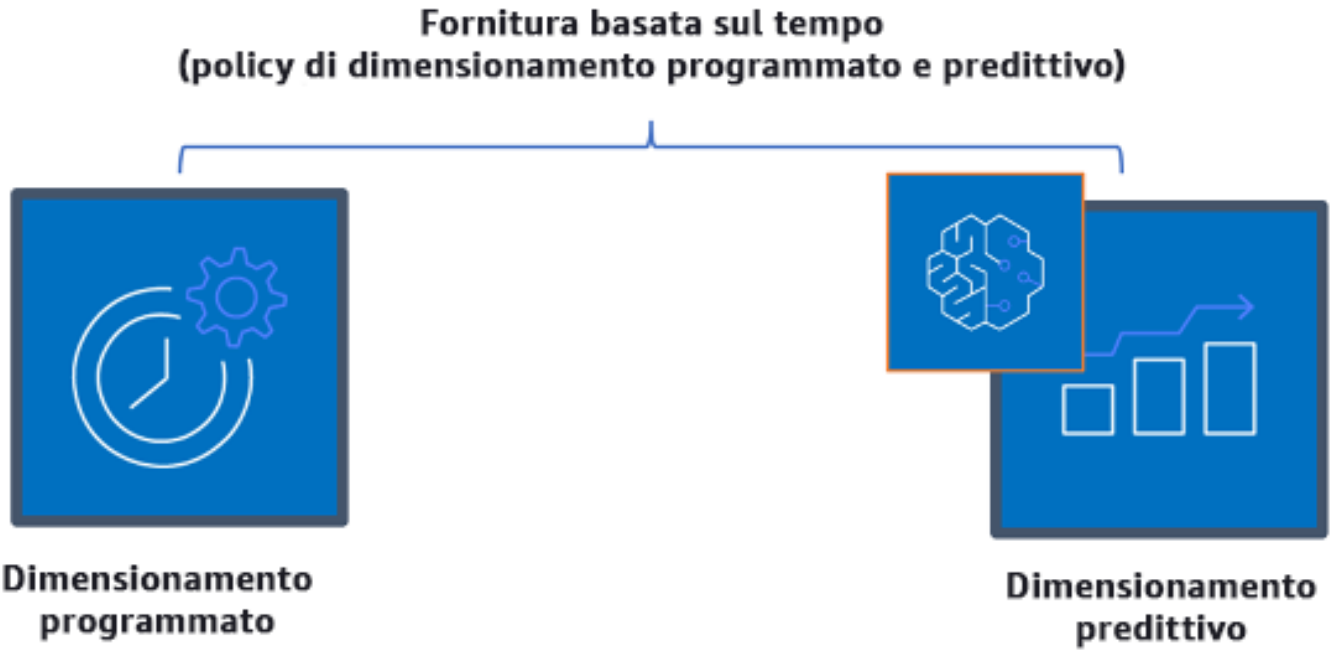
Policy di dimensionamento dinamico basato sulla domanda

- Dimensionamento semplice/graduale: monitora le metriche e aggiunge/rimuove le istanze secondo i passaggi definiti manualmente dai clienti.

- Monitoraggio degli obiettivi: meccanismo di controllo simile a un termostato che aggiunge o rimuove automaticamente le istanze per mantenere le metriche in base a un obiettivo definito dal cliente.

Quando prevedi una strategia basata sulla domanda in un progetto, tieni presenti due considerazioni principali. In primo luogo, devi capire con quale velocità è necessario predisporre le nuove risorse. In secondo luogo, devi capire che la dimensione del margine tra domanda e risorse fornite cambierà. Devi prepararti ad affrontare le variazioni nella domanda, nonché le risorse insufficienti.

Fornitura basata sul tempo: una strategia basata sul tempo allinea la capacità delle risorse alla domanda, che è prevedibile o ben definita nel tempo. In genere questa strategia non dipende dai livelli di utilizzo delle risorse. Una strategia basata sul tempo assicura che le risorse siano disponibili nel momento esatto in cui vengono richieste e possano essere fornite senza ritardi dovuti alle procedure di avvio e ai controlli di sistema o di coerenza. Attraverso una strategia basata sul tempo si possono fornire risorse aggiuntive o incrementare la capacità nei periodi più intensi.



Policy di dimensionamento basato sul tempo

Puoi utilizzare il dimensionamento automatico pianificato e predittivo per implementare un approccio basato sul tempo. I carichi di lavoro possono essere programmati per aumentare orizzontalmente in determinati momenti (ad esempio, all'inizio dell'orario di lavoro), garantendo quindi la disponibilità

delle risorse all'arrivo degli utenti on demand. Il dimensionamento predittivo utilizza modelli per aumentare orizzontalmente, mentre il dimensionamento pianificato utilizza tempi predefiniti per aumentare orizzontalmente. È inoltre possibile utilizzare la strategia di selezione del tipo di istanza (ABS) basata sugli attributi nei gruppi di Auto Scaling, che consente di esprimere i requisiti dell'istanza sotto forma di un insieme di attributi, come vCPU, memoria e archiviazione. Ciò consente inoltre di utilizzare automaticamente i tipi di istanze di nuova generazione quando vengono rilasciati e di accedere a una gamma più ampia di capacità con le istanze Amazon EC2 Spot. Amazon EC2 Fleet e Amazon EC2 Auto Scaling selezionano e avviano le istanze che soddisfano gli attributi specificati, eliminando la necessità di selezionare manualmente i tipi di istanza.

Puoi anche sfruttare e fornire AWS APIsSDKse AWS CloudFormationdisattivare automaticamente interi ambienti quando ne hai bisogno. Questa strategia risulta particolarmente adatta per gli ambienti di sviluppo o di prova che operano solo in determinati orari di lavoro o periodi di tempo. È possibile utilizzarlo APIs per scalare le dimensioni delle risorse all'interno di un ambiente (scalabilità verticale). Ad esempio, potresti aumentare verticalmente un carico di lavoro di produzione modificando la dimensione o la classe dell'istanza. Ciò è possibile interrompendo e avviando l'istanza e selezionando una dimensione o classe diversa. Questa tecnica può essere applicata anche ad altre risorse, come Amazon EBS Elastic Volumes, che possono essere modificate per aumentare le dimensioni, regolare le prestazioni (IOPS) o cambiare il tipo di volume durante l'uso.

Quando prevedi una strategia basata sul tempo in un progetto, tieni presenti due considerazioni principali. In primo luogo, che livello di coerenza presenta il modello di utilizzo? In secondo luogo, qual è l'impatto se il modello cambia? Puoi migliorare l'accuratezza delle previsioni monitorando i tuoi carichi di lavoro e utilizzando la business intelligence. Se si notano cambiamenti significativi nel modello di utilizzo, si possono modificare i tempi per assicurarti che la copertura sia fornita.

Passaggi dell'implementazione

- Configura il dimensionamento pianificato: per le variazioni prevedibili della domanda, il dimensionamento basato sul tempo può fornire il numero corretto di risorse in modo tempestivo. È utile anche se la creazione e la configurazione delle risorse non avvengono in maniera sufficientemente rapida per rispondere alle modifiche on demand. Utilizzando l'analisi del carico di lavoro, configura il dimensionamento pianificato utilizzando AWS Auto Scaling. Per configurare la pianificazione basata sul tempo, puoi utilizzare la scalabilità predittiva della scalabilità pianificata per aumentare in anticipo il numero di istanze Amazon EC2 nei tuoi gruppi di Auto Scaling in base alle variazioni di carico previste o prevedibili.
- Configura la scalabilità predittiva: la scalabilità predittiva ti consente di aumentare il numero di EC2 istanze Amazon nel tuo gruppo Auto Scaling in anticipo rispetto agli schemi giornalieri e settimanali

dei flussi di traffico. Se si hanno picchi di traffico regolari e applicazioni che richiedono molto tempo per avviarsi, si dovrebbe prendere in considerazione l'utilizzo del dimensionamento predittivo. Il dimensionamento predittivo può aiutare a scalare più velocemente inizializzando la capacità prima del carico previsto rispetto al solo dimensionamento dinamico, che è di natura reattiva. Ad esempio, se gli utenti iniziano a utilizzare il carico di lavoro all'inizio dell'orario di lavoro e non lo utilizzano dopo l'orario di lavoro, il dimensionamento predittivo può aggiungere capacità prima dell'orario di lavoro, eliminando i ritardi del dimensionamento dinamico per reagire alle variazioni del traffico.

- Configura il dimensionamento automatico dinamico: per configurare il dimensionamento in base ai parametri del carico di lavoro attivi, utilizza Auto Scaling. Utilizza l'analisi e configura Auto Scaling per l'avvio sui livelli di risorse corretti e assicurati che il carico di lavoro si riduca orizzontalmente nel tempo richiesto. È possibile avviare e scalare automaticamente un parco di istanze on demand e istanze spot all'interno di un singolo gruppo con un singolo gruppo Auto Scaling. Oltre a ricevere sconti per l'utilizzo di Istanze Spot, è possibile utilizzare Istanze riservate o Savings Plan per ricevere tariffe scontate sul normale prezzo delle istanze on demand. Tutti questi fattori combinati ti aiutano a ottimizzare i risparmi sui costi per EC2 le istanze Amazon e ti aiutano a ottenere la scalabilità e le prestazioni desiderate per la tua applicazione.

Risorse

Documenti correlati:

- AWS Auto Scaling
- AWS Instance Scheduler
- Dimensionare le dimensioni del gruppo con scalabilità automatica
- Guida introduttiva ad Amazon EC2 Auto Scaling
- Guida introduttiva ad Amazon SQS
- Scalabilità pianificata per Amazon EC2 Auto Scaling
- Scalabilità predittiva per Amazon EC2 Auto Scaling

Video correlati:

- Target Tracking Scaling Policies for Auto Scaling
- AWS Pianificatore di istanze

Esempi correlati:

- Selezione del tipo di istanza basata sugli attributi per Auto Scaling for Amazon Fleet EC2
- Optimizing Amazon Elastic Container Service for cost using scheduled scaling
- Scalabilità predittiva con Amazon EC2 Auto Scaling
- Come posso utilizzare Instance Scheduler con AWS CloudFormation per pianificare le EC2 istanze Amazon?

Ottimizzazione nel tempo

Questions

- COST10. In che modo valuti i nuovi servizi?
- COST11. Come valuti il costo dell'impegno?

COST10. In che modo valuti i nuovi servizi?

Non appena vengono AWS rilasciati nuovi servizi e funzionalità, è consigliabile rivedere le decisioni architettoniche esistenti per verificare che continuino a essere le più convenienti.

Best practice

- COST10-BP01 Sviluppare un processo di revisione del carico di lavoro
- COST10-BP02 Rivedi e analizza regolarmente questo carico di lavoro

COST10-BP01 Sviluppare un processo di revisione del carico di lavoro

Sviluppa un processo che definisca i criteri e il processo per la revisione del carico di lavoro. L'impegno analitico deve riflettere il potenziale risultato. Ad esempio, i carichi di lavoro principali o i carichi di lavoro con un valore superiore al 10% della fattura sono analizzati trimestralmente oppure ogni sei mesi, mentre i carichi di lavoro inferiori al 10% sono analizzati annualmente.

Livello di rischio associato se questa best practice non fosse adottata: elevato

Guida all'implementazione

Per far sì che il carico di lavoro sia sempre efficiente in termini di costi, devi analizzarlo regolarmente per stabilire se ci sono opportunità di implementare nuovi servizi, funzionalità e componenti. Per

garantire costi complessivi ridotti, il processo deve essere proporzionale al potenziale risparmio. Ad esempio, i carichi di lavoro che rappresentano il 50% della spesa complessiva devono essere esaminati con maggiore regolarità e più nel dettaglio rispetto ai carichi di lavoro che rappresentano il 5% della spesa complessiva. Prendi in considerazione qualsiasi fattore esterno o volatilità. Se il carico di lavoro serve una determinata area geografica o un segmento di mercato e viene previsto un cambiamento in tale area, revisioni più frequenti possono portare a risparmi sui costi. Un altro fattore in fase di revisione è rappresentato dall'impegno necessario per implementare le modifiche. Se i test e la convalida delle modifiche comportassero costi significativi, le revisioni dovrebbero essere meno frequenti.

Prendi in considerazione il costo nel lungo termine della manutenzione di componenti e risorse obsoleti e legacy, nonché dell'impossibilità di implementare in essi nuove funzionalità. L'attuale costo del test e della convalida potrebbe superare il vantaggio auspicato. Tuttavia, nel corso del tempo, il costo di apportare modifiche potrebbe crescere in modo significativo all'aumentare del divario tra il carico di lavoro e le tecnologie attuali, generando costi ancora maggiori. Ad esempio, il costo del passaggio a un nuovo linguaggio di programmazione potrebbe attualmente non risultare conveniente. Tuttavia, nel giro di cinque anni, il costo del personale qualificato per tale linguaggio potrebbe aumentare e, a causa dell'aumento del carico di lavoro, potresti dover trasferire un sistema ancora più grande al nuovo linguaggio, richiedendo sforzi ancora maggiori rispetto a prima.

Suddividi il carico di lavoro in componenti, assegna un costo ai componenti (una stima è sufficiente), quindi elenca i fattori (ad esempio, impegno richiesto e mercati esterni) accanto a ciascun componente. Utilizza questi indicatori per determinare una frequenza di revisione per ogni carico di lavoro. Ad esempio, potresti avere i server Web come un costo elevato, con un impegno di modifica ridotto e fattori esterni elevati, e da questo potrebbe derivare un'alta frequenza di revisione. Un database centrale può avere un costo medio, con un impegno di modifica elevato e un basso fattore esterno, e da questo potrebbe derivare una frequenza di revisione media.

Definisci un processo per valutare i nuovi servizi, i modelli di progettazione, i tipi di risorse e le configurazioni per ottimizzare il costo del tuo carico di lavoro man mano che diventano disponibili. Proprio come avviene nella revisione del pilastro delle prestazioni e nella revisione del pilastro dell'affidabilità, identifica, convalida e assegna la priorità alle attività di ottimizzazione e miglioramento e alla risoluzione dei problemi, quindi inseriscile nel tuo backlog.

Passaggi dell'implementazione

- Definisci la frequenza della revisione: definisci la frequenza con cui il carico di lavoro e i relativi componenti devono essere revisionati. Dedica tempo e risorse al miglioramento continuo e alla frequenza di revisione per migliorare l'efficienza e l'ottimizzazione del carico di lavoro. Si

tratta di una combinazione di fattori e può variare da carico di lavoro a carico di lavoro all'interno dell'organizzazione, ma può anche variare tra i componenti del carico di lavoro. I fattori più comuni sono: l'importanza per l'organizzazione misurata in termini di fatturato o marchio, il costo totale di esecuzione del carico di lavoro (inclusi costi operativi e delle risorse), la complessità del carico di lavoro, la facilità di implementazione di una modifica, eventuali accordi di licenza software e l'eventuale aumento dei costi di licenza dovuti a licenze punitive in seguito a una modifica. I componenti possono essere definiti a livello funzionale o tecnico come server Web e database, oppure come risorse di calcolo e storage. Equilibra i fattori di conseguenza e prevedi un periodo per il carico di lavoro e i relativi componenti. Puoi decidere di rivedere l'intero carico di lavoro ogni 18 mesi, esaminare i server Web ogni 6 mesi, il database ogni 12 mesi, l'elaborazione e lo storage a breve termine ogni 6 mesi e lo storage a lungo termine ogni 12 mesi.

- Definisci la completezza della revisione: stabilisci quanto impegno deve essere impiegato per la revisione dei componenti o dell'intero carico di lavoro. Come per la frequenza di revisione, si tratta di un equilibrio tra più fattori. Valuta e dai priorità alle opportunità di miglioramento per concentrare gli sforzi dove producono i vantaggi maggiori, stimando l'impegno necessario per queste attività. Se i risultati non sono in linea con gli obiettivi e l'impegno richiesto ha un costo superiore, riprova utilizzando linee d'azione alternative. I processi di revisione devono prevedere l'allocazione di tempo e risorse per rendere possibile il miglioramento incrementale continuo. Ad esempio, si può decidere di dedicare una settimana all'analisi del componente del database, una settimana di analisi alle risorse di calcolo e quattro ore alla revisione dell'archiviazione.

Risorse

Documenti correlati:

- AWS Blog di notizie
- Tipi di cloud computing
- Novità di AWS

Esempi correlati:

- AWS Support Proactive Services
- Revisioni periodiche dei carichi di lavoro SAP

COST10-BP02 Rivedi e analizza regolarmente questo carico di lavoro

I carichi di lavoro esistenti vengono rivisti con regolarità in base a ogni processo definito per scoprire se è possibile adottare nuovi servizi, se i servizi esistenti possono essere sostituiti o se i carichi di lavoro possono essere riprogettati.

Livello di rischio associato se questa best practice non fosse adottata: medio

Guida all'implementazione

AWS aggiunge costantemente nuove funzionalità in modo da poter sperimentare e innovare più velocemente con la tecnologia più recente. AWS What's New descrive in dettaglio come AWS si sta procedendo in tal senso e fornisce una rapida panoramica dei AWS servizi, delle funzionalità e degli annunci di espansione a livello regionale non appena vengono rilasciati. Puoi approfondire i rilasci previsti e usarli per la revisione e l'analisi dei tuoi carichi di lavoro esistenti. Per sfruttare i vantaggi dei nuovi AWS servizi e funzionalità, è necessario esaminare i carichi di lavoro e implementare nuovi servizi e funzionalità in base alle esigenze. Ciò significa che potrebbe essere necessario sostituire i servizi esistenti utilizzati per il carico di lavoro o modernizzare il carico di lavoro per adottare questi nuovi servizi. AWS Ad esempio, è possibile esaminare i carichi di lavoro e sostituire il componente di messaggistica con Amazon Simple Email Service. Ciò elimina il costo di gestione e manutenzione di un parco istanze, fornendo al contempo tutte le funzionalità a un costo ridotto.

Per analizzare il tuo carico di lavoro e individuare le opportunità potenziali, dovresti prendere in considerazione non solo i nuovi servizi, ma anche le nuove modalità per creare le soluzioni. Guarda i video di This is My Architecture su AWS per conoscere i progetti architettonici di altri clienti, le loro sfide e le loro soluzioni. Dai un'occhiata alla serie All-In per scoprire le applicazioni dei AWS servizi nel mondo reale e le storie dei clienti. Puoi inoltre guardare la serie di video Back to Basics che illustra, esamina e analizza le best practice di base relative ai modelli di architettura cloud. Un'altra fonte sono i video How to Build This, progettati per aiutare le persone con grandi idee su come dare vita al loro prodotto minimo redditizio (MVP) utilizzando AWS i servizi. È un modo per i costruttori di tutto il mondo che hanno una forte idea di ottenere indicazioni architettoniche da AWS Solutions Architects esperti. Infine, puoi consultare i materiali della risorsa Nozioni di base, che offre tutorial dettagliati.

Prima di avviare il processo di revisione segui i requisiti aziendali per il carico di lavoro, i requisiti sulla privacy dei dati e la sicurezza per usare un servizio o un'area geografica specifica e i requisiti di performance, seguendo al tempo stesso il processo di revisione concordato.

Passaggi dell'implementazione

- Rivedi con regolarità il carico di lavoro: utilizzando il processo definito, esegui le revisioni con la frequenza specificata. Accertati di dedicare la quantità di impegno necessaria per ciascun componente. Questo processo è simile a quello di progettazione iniziale in cui hai selezionato i servizi per l'ottimizzazione dei costi. Analizza i servizi e i vantaggi che porterebbero; questa volta considera anche il costo del tempo necessario per la modifica, non solo i vantaggi a lungo termine.
- Implementa nuovi servizi: se in seguito all'analisi ritieni di dover implementare modifiche, esegui innanzitutto una baseline del carico di lavoro per scoprire il costo corrente per ogni output. Implementa le modifiche, quindi esegui un'analisi per verificare il nuovo costo per ogni output.

Risorse

Documenti correlati:

- AWS Blog di notizie
- Novità di AWS
- AWS Documentazione
- AWS Guida introduttiva
- AWS Risorse generali

Video correlati:

- AWS - Questa è la mia architettura
- AWS - Ritorno alle basi
- AWS - Serie All-In
- How to Build This

COST11. Come valuti il costo dell'impegno?

Best practice

- COST11-BP01 Esegui l'automazione delle operazioni

COST11-BP01 Esegui l'automazione delle operazioni

Valuta i costi operativi del cloud, concentrandoti sulla quantificazione del risparmio di tempo e impegno nelle attività amministrative e nelle implementazioni, sulla mitigazione del rischio di errore

umano, sulla conformità e su altre operazioni tramite l'automazione. Valuta il tempo e i costi associati necessari per gli impegni operativi e implementa l'automazione per le attività amministrative per ridurre al minimo il lavoro manuale laddove possibile.

Livello di rischio associato se questa best practice non fosse adottata: basso

Guida all'implementazione

L'automazione delle operazioni riduce la frequenza di attività manuali, migliora l'efficienza e offre vantaggi ai clienti con un'esperienza affidabile e coerente durante l'implementazione, l'amministrazione o l'operatività dei carichi di lavoro. Puoi liberare le risorse dell'infrastruttura dalle attività operative manuali e usarle per operazioni e innovazioni di maggior valore, migliorando così i risultati aziendali. Le aziende vogliono un modo testato e collaudato di gestire i propri carichi di lavoro nel cloud. Tale soluzione deve essere sicura, veloce ed economica, con il minimo rischio e la massima affidabilità.

Inizia assegnando le priorità alle tue operazioni sulla base dell'impegno richiesto, considerando i costi complessivi. Ad esempio, quanto tempo è necessario per implementare nuove risorse nel cloud, eseguire modifiche di ottimizzazione alle risorse esistenti o implementare le configurazioni necessarie? Esamina il costo totale delle attività eseguite dal personale, tenendo conto dei costi operativi e di gestione. Dai la priorità alle automazioni per le attività amministrative per ridurre il livello di impegno delle persone.

L'impegno di revisione deve riflettere il potenziale risultato. Ad esempio, esamina il tempo impiegato per eseguire le attività manualmente rispetto a quello per eseguirle in automatico. Dai priorità all'automazione di attività ripetitive e di valore elevato che richiedono tempo e sono complesse. Le attività che presentano un rischio o un valore elevato di errore umano sono in genere il punto di partenza migliore da cui iniziare con l'automazione, poiché il rischio spesso comporta un costo operativo aggiuntivo indesiderato (come gli straordinari del team operativo).

Utilizza strumenti di automazione come AWS Systems Manager o AWS Config per semplificare le operazioni, la conformità, il monitoraggio, il ciclo di vita e i processi di terminazione. Con AWS servizi, strumenti e prodotti di terze parti, puoi personalizzare le automazioni implementate per soddisfare le tue esigenze specifiche. La tabella seguente mostra alcune delle funzioni e delle caratteristiche operative di base che puoi ottenere con i servizi AWS per automatizzare attività amministrative e operative:

- AWS Audit Manager: Verifica continuamente l' AWS utilizzo per semplificare la valutazione del rischio e della conformità

- AWS Backup: gestione e automazione centralizzata della protezione dei dati.
- AWS Config: configurazione delle risorse di elaborazione, valutazione, audit, esame delle configurazioni e dell'inventario delle risorse.
- AWS CloudFormation: avvio di risorse ad alta disponibilità con il modello Infrastructure as Code.
- AWS CloudTrail: gestione, conformità e controllo delle modifiche IT.
- Amazon EventBridge Pianifica eventi e attiva AWS Lambda l'azione.
- AWS Lambda: Automatizza i processi ripetitivi attivandoli con eventi o eseguendoli secondo una pianificazione fissa con. AWS EventBridge
- AWS Systems Manager: avvia e arresta i carichi di lavoro, applica patch ai sistemi operativi, automatizza la configurazione e la gestione continua.
- AWS Step Functions: pianificazione di lavori e automazione dei flussi di lavoro.
- AWS Service Catalog: utilizzo dei modelli, modello Infrastructure as code con conformità e controllo.

Se desideri adottare immediatamente le automazioni utilizzando AWS prodotti e servizi e non hai competenze nella tua organizzazione, contatta AWS Managed Services (AMS), i Servizi AWS professionali o AWS i partner per aumentare l'adozione dell'automazione e migliorare la tua eccellenza operativa nel cloud.

AWS Managed Services (AMS) è un servizio che gestisce AWS l'infrastruttura per conto di clienti e partner aziendali. Fornisce un ambiente sicuro e conforme in cui è possibile distribuire i carichi di lavoro. AMSutilizza modelli operativi cloud aziendali con automazione per consentire di soddisfare i requisiti dell'organizzazione, passare al cloud più rapidamente e ridurre i costi di gestione continui.

AWS I servizi professionali possono anche aiutarvi a raggiungere i risultati aziendali desiderati e ad automatizzare le operazioni con. AWS Consente ai clienti di implementare operazioni IT automatizzate, solide e agili, nonché funzionalità di governance ottimizzate per il cloud. Per esempi di monitoraggio dettagliati e best practice consigliate, consulta il whitepaper sul pilastro dell'eccellenza operativa.

Passaggi dell'implementazione

- Crea una sola volta e distribuisci più volte: usa infrastructure-as-code come CloudFormation AWS SDK, oppure AWS CLI distribuisci una sola volta e usa più volte per ambienti simili o per scenari di disaster recovery. Applica i tag durante l'implementazione per monitorare il tuo consumo definito

in altre best practice. Utilizzalo AWS Launch Wizardper ridurre i tempi di implementazione di molti carichi di lavoro aziendali più diffusi. AWS Launch Wizard ti guida attraverso il dimensionamento, la configurazione e la distribuzione dei carichi di lavoro aziendali seguendo le best practice. AWS Puoi anche utilizzare Service Catalog, che ti aiuta a creare e gestire modelli infrastructure-as-code approvati da utilizzare in AWS modo che chiunque possa scoprire risorse cloud self-service approvate.

- Automatizza la conformità continua: prendi in considerazione l'automazione di valutazioni e correzioni delle configurazioni registrate rispetto agli standard predefiniti. Se si AWS Organizations combinano le funzionalità di AWS Config and AWS CloudFormation, è possibile gestire e automatizzare in modo efficiente la conformità alla configurazione su larga scala per centinaia di account membri. È possibile esaminare le modifiche nelle configurazioni e nelle relazioni tra le AWS risorse e approfondire la cronologia di una configurazione delle risorse.
- Automatizza le attività di monitoraggio: AWS offre svariati strumenti di monitoraggio dei servizi. Puoi configurare questi strumenti per automatizzare le attività di monitoraggio. Crea e implementa un piano di monitoraggio che raccolga i dati da tutte le parti del carico di lavoro in modo da poter eseguire più facilmente il debug di un errore su più punti, se si verifica. Ad esempio, puoi utilizzare gli strumenti di monitoraggio automatizzato per osservare Amazon EC2 e segnalarti quando qualcosa non va per i controlli dello stato del sistema, i controlli dello stato delle istanze e gli CloudWatch allarmi Amazon.
- Automatizza manutenzione e operazioni: esegui in automatico operazioni di routine senza l'intervento umano. Utilizzando AWS servizi e strumenti, puoi scegliere quali AWS automazioni implementare e personalizzare in base alle tue esigenze specifiche. Ad esempio, utilizza EC2Image Builder per creare, testare e distribuire immagini di macchine virtuali e container da utilizzare in locale AWS o per applicare patch alle istanze. EC2 AWS SSM Se l'azione desiderata non può essere eseguita con AWS i servizi o hai bisogno di azioni più complesse con il filtraggio delle risorse, automatizza le operazioni utilizzando () o gli strumenti. AWS Command Line InterfaceAWS CLI AWS SDK AWS CLI offre la possibilità di automatizzare l'intero processo di controllo e gestione dei AWS servizi con script senza utilizzare il. AWS Management Console Seleziona il tuo preferito AWS SDKs per interagire con AWS i servizi. Per altri esempi di codice, consulta AWS SDK Code examples repository.
- Crea un ciclo di vita continuo con le automazioni: è importante stabilire e preservare policy consolidate del ciclo di vita non solo per le normative o la ridondanza, ma anche per l'ottimizzazione dei costi È possibile AWS Backup utilizzarlo per gestire e automatizzare centralmente la protezione dei dati degli archivi di dati, come bucket, volumi, database e file system. Puoi anche utilizzare Amazon Data Lifecycle Manager per automatizzare la creazione, la conservazione e l'eliminazione di snapshot e backup. EBS EBS AMIs

- Eliminare risorse non necessarie: è abbastanza comune accumulare risorse inutilizzate nella sandbox o nello sviluppo. Account AWS Gli sviluppatori creano e sperimentano vari servizi e risorse come parte del normale ciclo di sviluppo, quindi non eliminano le risorse quando non sono più necessarie. Le risorse inutilizzate possono comportare costi superflui e talvolta elevati per l'organizzazione. L'eliminazione di queste risorse può ridurre i costi operativi di questi ambienti. Assicurati che i dati non siano necessari o esegui un backup se non sei sicuro. È possibile usare AWS CloudFormation per pulire gli stack implementati, eliminando automaticamente la maggior parte delle risorse definite nel modello. In alternativa, puoi creare un'automazione per l'eliminazione delle AWS risorse utilizzando strumenti come aws-nuke.

Risorse

Documenti correlati:

- Modernizzazione delle operazioni in Cloud AWS
- AWS Services for Automation
- Infrastructure and automation
- AWS Systems Manager Automation
- Monitoraggio automatico e manuale
- AWS automazioni per l'SAPamministrazione e le operazioni
- AWS Managed Services
- AWS Professional Services

Video correlati:

- Automatizza la conformità continua su larga scala in AWS
- AWS Backup Demo: Backup su più account e più regioni
- Applicazione di patch per le tue istanze Amazon EC2

Esempi correlati:

- Reinventing automated operations (Part I)
- Reinventing automated operations (Part II)
- Automatizza l'eliminazione delle AWS risorse utilizzando aws-nuke

- Elimina i EBS volumi Amazon non utilizzati utilizzando AWS Config e AWS SSM
- Automatizza la conformità continua su larga scala in AWS
- Automazioni IT con AWS Lambda

Sostenibilità

Il pilastro della sostenibilità include la consapevolezza dell'impatto dei servizi utilizzati, la quantificazione di tale impatto per l'intero ciclo di vita del carico di lavoro e l'applicazione dei principi di progettazione e delle best practice per ridurlo nella fase di sviluppo di carichi di lavoro cloud. Puoi trovare linee guida prescrittive sull'implementazione nel whitepaper sul pilastro della sostenibilità.

Aree delle best practice

- Selezione della regione
- Allineamento alla domanda
- Software e architettura
- Dati
- Hardware e servizi
- Processo e cultura

Selezione della regione

Domanda

- SUS1 Come selezionate le regioni per il vostro carico di lavoro?

SUS1 Come selezionate le regioni per il vostro carico di lavoro?

La scelta della regione per il carico di lavoro influisce in modo significativo su prestazioniKPIs, costi e impronta di carbonio. Per migliorarli efficacementeKPIs, dovresti scegliere le regioni per i tuoi carichi di lavoro in base ai requisiti aziendali e agli obiettivi di sostenibilità.

Best practice

- SUS01-BP01 Scegli la regione in base ai requisiti aziendali e agli obiettivi di sostenibilità

SUS01-BP01 Scegli la regione in base ai requisiti aziendali e agli obiettivi di sostenibilità

Scegli una regione per il tuo carico di lavoro in base ai tuoi requisiti aziendali e agli obiettivi di sostenibilità per ottimizzarloKPIs, inclusi prestazioni, costi e impronta di carbonio.

Anti-pattern comuni:

- Selezione della regione del carico di lavoro in base alla propria collocazione.
- Consolidamento di tutte le risorse del carico di lavoro in un'unica posizione geografica.

Vantaggi dell'adozione di questa best practice: riduzione dell'impronta di carbonio di un carico di lavoro collocandolo vicino ai progetti legati alle energie rinnovabili di Amazon o alle regioni con un'intensità ridotta di emissione di anidride carbonica.

Livello di rischio associato se questa best practice non fosse adottata: medio

Guida all'implementazione

Cloud AWS Si tratta di una rete in continua espansione di regioni e punti di presenza (PoP), con un'infrastruttura di rete globale che li collega tra loro. La scelta della regione per il carico di lavoro influisce in modo significativo su prestazioniKPIs, costi e impronta di carbonio. Per migliorarli efficacementeKPIs, dovresti scegliere le regioni per il tuo carico di lavoro in base ai tuoi requisiti aziendali e agli obiettivi di sostenibilità.

Passaggi dell'implementazione

- Segui questi passaggi per valutare e selezionare le potenziali regioni per il tuo carico di lavoro in base ai requisiti aziendali, tra cui la conformità, le funzionalità disponibili, il costo e la latenza.
  - Conferma che queste regioni siano conformi in base alle normative locali richieste.
  - Consulta gli elenchi dei servizi AWS per regione per verificare la presenza nelle regioni di servizi e funzionalità adeguati alla gestione del tuo carico di lavoro.
  - Calcola il costo del carico di lavoro per ciascuna regione mediante il AWS Pricing Calculator.
  - Verifica la latenza di rete tra le diverse sedi degli utenti finali. Regione AWS
- Scegli le regioni in prossimità dei progetti di generazione di energia rinnovabile di Amazon e le regioni in cui la griglia presenta un'intensità di emissione di anidride carbonica nota inferiore a quella di altre sedi (o regioni).

- Identifica le linee guida di sostenibilità pertinenti per tracciare e confrontare le emissioni di year-to-year carbonio sulla base del Greenhouse Gas Protocol (metodi basati sul mercato e basati sulla posizione).
- Scegli la regione in base al metodo utilizzato per monitorare le emissioni di anidride carbonica. Per ulteriori informazioni circa la scelta di una regione in base alle tue linee guida sulla sostenibilità, consulta How to select a Region for your workload based on sustainability goals.

Risorse

Documenti correlati:

- Understanding your carbon emission estimations
- Amazon Around the Globe
- Renewable Energy Methodology
- What to Consider when Selecting a Region for your Workloads

Video correlati:

- AWS re:Invent 2023 - Innovazione sostenibile nelle infrastrutture globali AWS
- AWS re:Invent 2023 - Architettura sostenibile: passato, presente e futuro
- AWS re:Invent 2022 - Fornire architetture sostenibili e ad alte prestazioni
- AWS re:Invent 2022 - Progettare in modo sostenibile e ridurre l'impronta di carbonio AWS
- AWS re:Invent 2022 - Sostenibilità nelle infrastrutture globali AWS

Allineamento alla domanda

Domanda

- SUS2 Come allineate le risorse cloud alla vostra richiesta?

SUS2 Come allineate le risorse cloud alla vostra richiesta?

Il modo in cui gli utenti e le applicazioni utilizzano i tuoi carichi di lavoro e altre risorse può aiutarti a identificare i miglioramenti da implementare per raggiungere gli obiettivi di sostenibilità. Puoi scalare l'infrastruttura in modo che sia costantemente adatta alla domanda e verifica di usare solo le risorse minime necessarie per supportare gli utenti. Allinea i livelli di servizio alle esigenze dei clienti. Colloca

le risorse in modo da limitare la rete necessaria per il loro consumo da parte di utenti e applicazioni. Rimuovi gli asset inutilizzati. Offri ai membri del team dispositivi in grado di soddisfarne le esigenze con un impatto minimo in termini di sostenibilità.

Best practice

- SUS02-BP01 Scala l'infrastruttura dei carichi di lavoro in modo dinamico
- SUS02-BP02 Allinearsi agli obiettivi di sostenibilità SLAs
- SUS02-BP03 Blocca la creazione e la manutenzione di risorse inutilizzate
- SUS02-BP04 Ottimizza il posizionamento geografico dei carichi di lavoro in base ai requisiti di rete
- SUS02-BP05 Ottimizza le risorse dei membri del team per le attività eseguite
- SUS02-BP06 Implementare il buffering o il throttling per appiattire la curva di domanda

SUS02-BP01 Scala l'infrastruttura dei carichi di lavoro in modo dinamico

Usa l'elasticità del cloud e dimensiona la tua infrastruttura in modo dinamico per rispondere alla richiesta di fornitura di risorse cloud ed evitare il provisioning eccessivo nel tuo carico di lavoro.

Anti-pattern comuni:

- Mancato dimensionamento dell'infrastruttura in base al carico degli utenti.
- Costante dimensionamento manuale dell'infrastruttura.
- Dopo un evento di dimensionamento, lasci una capacità aumentata anziché ridurre il dimensionamento.

Vantaggi dell'adozione di questa best practice: configurazione e test dell'elasticità del carico di lavoro consentono di abbinare in modo ottimale l'offerta di risorse cloud alla domanda ed evitare capacità con un provisioning eccessivo. Puoi sfruttare i vantaggi dell'elasticità nel cloud per scalare automaticamente la capacità durante e dopo i picchi di richiesta ed essere sicuro di utilizzare solo il numero esatto di risorse necessario per soddisfare le esigenze aziendali.

Livello di rischio associato se questa best practice non fosse adottata: medio

Guida all'implementazione

Il cloud offre la flessibilità necessaria per espandere o ridurre le risorse in modo dinamico attraverso una serie di meccanismi per soddisfare i cambiamenti della domanda. La corrispondenza ottimale tra offerta e domanda consente l'impatto ambientale più basso per un carico di lavoro.

La domanda può essere fissa o variabile e richiede parametri e automazione, allo scopo di garantire che la gestione non diventi particolarmente onerosa. Le applicazioni possono essere scalate verticalmente (verso l'alto o verso il basso) modificando la dimensione dell'istanza, orizzontalmente (aumentando o diminuendo) modificando il numero di istanze o tramite una combinazione delle due opzioni.

Puoi adottare varie strategie di approccio per associare l'offerta di risorse alla domanda.

- Approccio al tracciamento degli obiettivi: monitora il parametro di dimensionamento e aumenta o diminuisci automaticamente la capacità in base alle esigenze.
- Dimensionamento predittivo: procedi a ridurre orizzontalmente in previsione delle tendenze giornaliere e settimanali.
- Approccio basato sulla pianificazione: imposta il tuo programma di dimensionamento in base alle variazioni di carico prevedibili.
- Scalabilità dei servizi: scegli servizi (come il serverless) dotati dimensionamento nativo per progettazione o con dimensionamento automatico come funzionalità.

Identifica i periodi di utilizzo assente o ridotto e dimensiona le risorse per evitare capacità in eccesso e migliorare il livello di efficienza.

Passaggi dell'implementazione

- L'elasticità corrisponde all'offerta di risorse disponibili rispetto alla relativa domanda. Istanze, contenitori e funzioni forniscono meccanismi di elasticità, in combinazione con la scalabilità automatica o come funzionalità del servizio. AWS offre una gamma di meccanismi di ridimensionamento automatico per garantire che i carichi di lavoro possano essere ridotti rapidamente e facilmente durante i periodi di basso carico degli utenti. Ecco alcuni esempi di meccanismi di dimensionamento automatico:

Meccanismo di dimensionamento automatico	Dove usarlo
Amazon EC2 Auto Scaling	Utilizzalo per verificare di avere il numero corretto di EC2 istanze Amazon disponibili per gestire il carico di utenti per la tua applicazi one.

Meccanismo di dimensionamento automatico	Dove usarlo
Application Auto Scaling	Utilizzalo per scalare automaticamente le risorse per singoli AWS servizi oltre AmazonEC2, come le funzioni Lambda o i servizi Amazon Elastic Container Service ECS (Amazon).
Kubernetes Cluster Autoscaler	Usalo per scalare automaticamente i cluster Kubernetes. AWS

- La scalabilità viene spesso discussa in relazione a servizi di elaborazione come EC2 le istanze o le funzioni di Amazon. AWS Lambda Prendi in considerazione la configurazione di servizi non di calcolo, come le unità di capacità di lettura e scrittura di Amazon DynamoDB o le partizioni del flusso di dati Amazon Kinesis per soddisfare la domanda.
- Verifica che le metriche per l'aumento verticale o orizzontale siano convalidate in base al tipo di carico di lavoro implementato. Se stai implementando un'applicazione di transcodifica video, è previsto un CPU utilizzo del 100% e non dovrebbe essere la metrica principale. Se necessario, puoi servirti di una metrica personalizzata (ad esempio, l'utilizzo della memoria) per la policy di dimensionamento. Per scegliere le metriche giuste, consulta le seguenti linee guida per AmazonEC2:
  - La metrica deve essere una metrica di utilizzo valida e descrivere il livello di impiego di un'istanza.
  - Il valore del parametro deve aumentare e diminuire in proporzione al numero di istanze nel gruppo con scalabilità automatica.
- Usa il dimensionamento dinamico anziché il dimensionamento manuale per il tuo gruppo Auto Scaling. È consigliabile utilizzare le policy di dimensionamento del monitoraggio degli obiettivi nel dimensionamento dinamico
- Verifica che le implementazioni dei carichi di lavoro siano in grado di aumentare orizzontalmente e ridurre orizzontalmente. Crea scenari di test per eventi in cui si procede a ridurre orizzontalmente per verificare che il carico di lavoro si comporti secondo le aspettative e che non incida sull'esperienza utente (come nel caso della perdita di sessioni persistenti). Ad esempio, puoi usare la cronologia delle attività per verificare le attività di dimensionamento per un gruppo Auto Scaling.
- Analizza il tuo carico di lavoro per individuare modelli prevedibili e dimensionare le tue risorse in modo proattivo, anticipando variazioni nella domanda previste e pianificate. Con il

dimensionamento predittivo puoi eliminare la necessità di offrire capacità in eccedenza. Per maggiori dettagli, consulta Predictive Scaling with Amazon Auto EC2 Scaling.

Risorse

Documenti correlati:

- Guida introduttiva ad Amazon EC2 Auto Scaling
- Scalabilità predittiva perEC2, basata sul Machine Learning
- Analizza il comportamento degli utenti utilizzando Amazon OpenSearch Service, Amazon Data Firehose e Kibana
- Che cos'è Amazon CloudWatch?
- Monitoraggio del carico del DB con Performance Insights su Amazon RDS
- Presentazione del supporto nativo per la scalabilità predittiva con Amazon EC2 Auto Scaling
- Introducing Karpenter - An Open-Source, High-Performance Kubernetes Cluster Autoscaler
- Approfondimento su Amazon ECS Cluster Auto Scaling

Video correlati:

- AWS re:Invent 2023: scalabile AWS per i primi 10 milioni di utenti
- AWS re:Invent 2023 - Architettura sostenibile: passato, presente e futuro
- AWS re:Invent 2022 - Crea un ambiente di elaborazione efficiente in termini di costi, energia e risorse
- AWS re:Invent 2022 - Scalabilità dei contenitori da un utente a milioni
- AWS re:Invent 2023 - Ridimensionamento dell'inferenza FM a centinaia di modelli con Amazon SageMaker
- AWS re:Invent 2023 - Sfrutta la potenza di Karpenter per scalare, ottimizzare e aggiornare Kubernetes

Esempi correlati:

- Autoscaling

SUS02-BP02 Allinearsi agli obiettivi di sostenibilità SLAs

Rivedi e ottimizza gli accordi sui livelli di servizio del carico di lavoro (SLA) in base agli obiettivi di sostenibilità per ridurre al minimo le risorse necessarie per supportare il carico di lavoro continuando a soddisfare le esigenze aziendali.

Anti-pattern comuni:

- I carichi di lavoro SLAs sono sconosciuti o ambigui.
- Sei tu a definire i tuoi SLA obiettivi in termini di disponibilità e prestazioni.
- Usi lo stello modello di progettazione (come l'architettura multi-AZ) per tutti i carichi di lavoro.

Vantaggi derivanti dall'adozione di questa best practice: l'allineamento SLAs agli obiettivi di sostenibilità porta a un utilizzo ottimale delle risorse soddisfacendo al contempo le esigenze aziendali.

Livello di rischio associato se questa best practice non fosse adottata: basso

Guida all'implementazione

SLAsdefinisci il livello di servizio previsto da un carico di lavoro cloud, ad esempio tempi di risposta, disponibilità e conservazione dei dati. Questi influenzano l'architettura, l'utilizzo delle risorse e l'impatto ambientale di un carico di lavoro nel cloud. A cadenza regolare, rivedi SLAs e fai dei compromessi che riducano in modo significativo l'utilizzo delle risorse in cambio di riduzioni accettabili dei livelli di servizio.

Passaggi dell'implementazione

- Analizza gli obiettivi di sostenibilità: individua gli obiettivi di sostenibilità della tua organizzazione, come la riduzione delle emissioni di carbonio o l'ottimizzazione dell'utilizzo delle risorse.
- RevisioneSLAs: valuta le tue SLAs per valutare se soddisfano i tuoi requisiti aziendali. Se stai superando i limitiSLAs, esegui un'ulteriore revisione.
- Analizza i compromessi: esamina i compromessi in termini di complessità del carico di lavoro (come un elevato volume di utenti simultanei), prestazioni (come la latenza) e impatto sulla sostenibilità (come le risorse richieste). In genere, dare la priorità a due fattori va a scapito del terzo.
- Adeguamento SLAs: aggiusta la SLAs situazione adottando compromessi che riducano in modo significativo gli impatti sulla sostenibilità in cambio di riduzioni accettabili dei livelli di servizio.

  - Sostenibilità e affidabilità: i carichi di lavoro a elevata disponibilità presentano la tendenza a un maggiore consumo di risorse.
  - Sostenibilità e prestazioni: l'utilizzo di più risorse per aumentare le prestazioni potrebbe tradursi in un maggiore impatto ambientale.
  - Sostenibilità e sicurezza: carichi di lavoro eccessivamente sicuri potrebbero avere un impatto ambientale maggiore.
- Definisci la sostenibilità, SLAs se possibile: includi la sostenibilità nel tuo carico di SLAs lavoro. Ad esempio, definisci un livello minimo di utilizzo come sostenibilità SLA per le tue istanze di calcolo.
- Utilizza modelli di progettazione efficienti: utilizza modelli di progettazione come i microservizi per dare priorità alle AWS funzioni aziendali critiche e consentire livelli di servizio inferiori (come obiettivi in termini di tempi di risposta o tempi di ripristino) per funzioni non critiche.
- Comunica e stabilisci la responsabilità: condividi le informazioni SLAs con tutte le parti interessate, inclusi il team di sviluppo e i clienti. Utilizza i report per tracciare e monitorare ilSLAs. Assegna la responsabilità per raggiungere i tuoi obiettivi di sostenibilità. SLAs
- Utilizza incentivi e premi: utilizza incentivi e premi per raggiungere o superare SLAs gli obiettivi di sostenibilità in linea con gli obiettivi di sostenibilità.
- Revisione e iterazione: rivedi e modifica regolarmente i tuoi obiettivi SLAs per assicurarti che siano in linea con l'evoluzione degli obiettivi di sostenibilità e prestazioni.

Risorse

Documenti correlati:

- Understand resiliency patterns and trade-offs to architect efficiently in the cloud
- Importance of Service Level Agreement for SaaS Providers

Video correlati:

- AWS re:Invent 2023 - Capacità, disponibilità, efficienza dei costi: scegline tre
- AWS re:Invent 2023 - Architettura sostenibile: passato, presente e futuro
- AWS re:Invent 2023 - Modelli di integrazione avanzati e compromessi per sistemi liberamente accoppiati
- AWS re:Invent 2022 - Fornire architetture sostenibili e ad alte prestazioni

- AWS re:Invent 2022 - Crea un ambiente di elaborazione efficiente in termini di costi, energia e risorse

SUS02-BP03 Blocca la creazione e la manutenzione di risorse inutilizzate

Disattiva le risorse non utilizzate nel tuo carico di lavoro per ridurre il numero di risorse cloud richieste per supportare la domanda e per ridurre gli sprechi.

Anti-pattern comuni:

- Non analizzi la tua applicazione per individuare le risorse ridondanti o non più necessarie.
- Non rimuovi le risorse ridondanti o non più necessarie.

Vantaggi dell'adozione di questa best practice: la rimozione delle risorse non utilizzati libera risorse e migliora l'efficienza complessiva del carico di lavoro cloud.

Livello di rischio associato se questa best practice non fosse adottata: basso

Guida all'implementazione

Le risorse inutilizzate consumano risorse cloud come spazio di archiviazione e potenza di elaborazione. Individuando ed eliminando queste risorse, puoi liberare capacità e ottenere un'architettura cloud più efficiente. Analizza le risorse delle applicazioni con regolarità (come report precompilati, set di dati, immagini statiche e modelli di accesso alle risorse) per identificare ridondanze, sottoutilizzi e obiettivi potenziali di disattivazione. Elimina le risorse ridondanti per ridurre gli sprechi nel tuo carico di lavoro.

Passaggi dell'implementazione

- Predisponi un inventario: redigi un inventario completo al fine di individuare tutte le risorse all'interno del tuo carico di lavoro.
- Analizza l'utilizzo: usa strumenti di monitoraggio per identificare risorse statiche non più necessarie.
- Rimuovi le risorse inutilizzate: predisponi un piano per la rimozione delle risorse non più necessarie.
  - Prima di rimuovere qualsiasi risorsa, valuta l'impatto della rimozione sull'architettura.
  - Analizza le risorse generate in sovrapposizione per rimuovere le elaborazioni ridondanti.
  - Aggiorna le tue applicazioni per smettere di produrre e archiviare risorse che non sono più necessarie.

- Comunica con le terze parti: indica alle terze parti di smettere di produrre e di archiviare per tuo conto risorse gestite non più necessarie. Chiedi di consolidare le risorse ridondanti.
- Usa le policy del ciclo di vita: serviti delle policy del ciclo di vita per l'eliminazione in automatico le risorse inutilizzate.
  - Puoi utilizzare il ciclo di vita Amazon S3 per gestire gli oggetti durante il loro ciclo di vita.
  - Puoi utilizzare Amazon Data Lifecycle Manager per automatizzare la creazione, la conservazione e l'eliminazione di istantanee EBS Amazon e Amazon -backed. EBS AMIs
- Rivedi e ottimizza: esamina con regolarità il tuo carico di lavoro per individuare e rimuovere risorse non utilizzate.

Risorse

Documenti correlati:

- Ottimizzazione dell' AWS infrastruttura per la sostenibilità, parte II: storage
- Come faccio a eliminare le risorse attive che non mi servono più sul mio computer? Account AWS

Video correlati:

- AWS re:Invent 2023 - Architettura sostenibile: passato, presente e futuro
- AWS re:Invent 2022 - Preservazione e ottimizzazione del valore delle risorse multimediali digitali con Amazon S3
- AWS re:Invent 2023 - Ottimizza i costi nei tuoi ambienti con più account

SUS02-BP04 Ottimizza il posizionamento geografico dei carichi di lavoro in base ai requisiti di rete

Seleziona le sedi cloud e i servizi per il carico di lavoro per ridurre la distanza che il traffico di rete deve percorrere e diminuire così le risorse totali di rete richieste per supportare il carico di lavoro.

Anti-pattern comuni:

- Selezione della regione del carico di lavoro in base alla propria collocazione.
- Consolidamento di tutte le risorse del carico di lavoro in un'unica posizione geografica.
- Tutto il traffico passa attraverso i data center esistenti.

Vantaggi dell'adozione di questa best practice: il posizionamento di un carico di lavoro in prossimità dei relativi utenti garantisce la latenza più bassa possibile e la contemporanea riduzione del trasferimento dei dati nella rete e dell'impatto ambientale.

Livello di rischio associato se questa best practice non fosse adottata: medio

Guida all'implementazione

L' Cloud AWS infrastruttura è costruita attorno a opzioni di localizzazione come Regioni, Zone di disponibilità, gruppi di collocamento e edge location come AWS OutpostsAWSLocal Zones. Queste opzioni relative alle sedi sono responsabili della gestione della connettività tra i componenti delle applicazioni, i servizi cloud, le reti edge e i data center on-premises.

Analizza i modelli di accesso alla rete nel tuo carico di lavoro per stabilire come usare queste opzioni relative alle sedi cloud e ridurre la distanza che il traffico di rete deve percorrere.

Passaggi dell'implementazione

- Analizza i modelli di accesso alla rete nel tuo carico di lavoro per capire come gli utenti usano la tua applicazione.
  - Utilizza strumenti di monitoraggio, come Amazon CloudWatch e AWS CloudTrail, per raccogliere dati sulle attività di rete.
  - Analizza i dati per identificare il modello di accesso alla rete.
- Seleziona le regioni appropriate per l'implementazione del carico di lavoro in base ai seguenti elementi chiave:
  - Il tuo obiettivo di sostenibilità: come illustrato nella sezione Selezione della regione.
  - Ubicazione dei dati per le applicazioni a uso intensivo di dati, ad esempio applicazioni di big data e machine learning, il codice dell'applicazione dovrebbe essere eseguito il più vicino possibile ai dati.
  - Ubicazione degli utenti: per le applicazioni rivolte agli utenti, scegli una regione o più regioni vicine agli utenti del carico di lavoro.
  - Altri vincoli: prendi in considerazione vincoli, come costi e conformità, come illustrato in What to Consider when Selecting a Region for your Workloads.
- Usa la cache locale o le soluzioni di caching AWS per i dati di frequente utilizzo per migliorare le performance, ridurre lo spostamento dei dati e minimizzare l'impatto ambientale.

Servizio	Quando usare
Amazon CloudFront	Utilizzalo per memorizzare nella cache contenuti statici come immagini, script e video, nonché contenuti dinamici come API risposte o applicazioni web.
Amazon ElastiCache	Usalo per memorizzare nella cache i contenuti per le applicazioni Web.
DynamoDB Accelerator	Usalo per aggiungere accelerazione in memoria alle tabelle DynamoDB.

- Utilizza servizi in grado di supportarti nell'esecuzione del codice in posizioni più vicine agli utenti del carico di lavoro:

Servizio	Quando usare
Lambda@Edge	Usalo per operazioni a uso intensivo di risorse di calcolo eseguite quando gli oggetti non si trovano nella cache.
CloudFront Funzioni Amazon	Utilizzalo per casi d'uso semplici come HTTP manipolazioni di richieste o risposte che possono essere avviate da funzioni di breve durata.
AWS IoT Greengrass	Usale per eseguire la memorizzazione nella cache di risorse di calcolo, messaggistica e dati per i dispositivi connessi.

- Utilizza il pooling delle connessioni per consentire il loro riutilizzo e ridurre le risorse richieste.
- Utilizza archivi di dati distribuiti che non si affidano a connessioni persistenti e aggiornamenti sincroni per garantire coerenza e servire le popolazioni regionali.
- Sostituisci la capacità di rete statica preallocata con una capacità dinamica condivisa e condividi l'impatto in termini di sostenibilità della capacità di rete con altri abbonati.

Risorse

Documenti correlati:

- Ottimizzazione dell' AWS infrastruttura per la sostenibilità, parte: rete III
- ElastiCache Documentazione Amazon
- Che cos'è Amazon CloudFront?
- Caratteristiche CloudFront principali di Amazon
- AWS Infrastruttura globale
- AWS Local Zones e AWS Outposts scelta della tecnologia giusta per il tuo carico di lavoro edge
- Placement groups
- AWS Local Zones
- AWS Outposts

Video correlati:

- Demistificazione del trasferimento di dati su AWS
- Scalabilità delle prestazioni di rete sulle istanze Amazon di nuova generazione EC2
- AWS Video esplicativo su Local Zones
- AWS Outposts: Overview and How it Works
- AWS re:Invent 2023 - Una strategia di migrazione per carichi di lavoro edge e locali
- AWS re:Invent 2021 -: Portare l'esperienza in sede AWS OutpostsAWS
- AWS re:Invent 2020 - AWS Wavelength: Esegui app con latenza ultra bassa sull'edge 5G
- AWS re:Invent 2022 - AWS Local Zones: creazione di applicazioni per un edge distribuito
- AWS re:Invent 2021 - Creazione di siti Web a bassa latenza con Amazon CloudFront
- AWS re:Invent 2022 - Migliora le prestazioni e la disponibilità con AWS Global Accelerator
- AWS re:Invent 2022 - Costruisci la tua rete WAN utilizzando AWS
- AWS re:Invent 2020: gestione globale del traffico con Amazon Route 53

Esempi correlati:

- AWS Workshop di networking
- Architecting for sustainability - Minimize data movement across networks

SUS02-BP05 Ottimizza le risorse dei membri del team per le attività eseguite

Ottimizza le risorse fornite ai membri del team per ridurre al minimo l'impatto sulla sostenibilità ambientale e supportare al tempo stesso le loro esigenze.

Anti-pattern comuni:

- Ignori l'impatto dei dispositivi utilizzati dai membri del tuo team sull'efficienza complessiva della tua applicazione cloud.
- Gestisci e aggiorni manualmente le risorse utilizzate dai membri del tuo team.

Vantaggi dell'adozione di questa best practice: migliore efficienza complessiva delle applicazioni abilitate per il cloud grazie all'ottimizzazione delle risorse dei membri del team.

Livello di rischio associato se questa best practice non fosse adottata: basso

Guida all'implementazione

Identifica le risorse che i membri del tuo team usano per accedere ai tuoi servizi, il loro ciclo di vita atteso e l'impatto finanziario e di sostenibilità. Implementa strategie per ottimizzare queste risorse. Esegui ad esempio operazioni complesse, come rendering e compilazione, su infrastrutture scalabili altamente utilizzate, invece che su sistemi per utenti singoli, sottoutilizzati e con un alto dispendio energetico.

Passaggi dell'implementazione

- Utilizza workstation efficienti dal punto di vista energetico: fornisci ai membri del team workstation e periferiche efficienti dal punto di vista energetico. Utilizza in questi dispositivi funzionalità di gestione dell'alimentazione efficienti, come la modalità di risparmio energetico, per ridurne il consumo di energia.
- Usa la virtualizzazione: usa desktop virtuali e lo streaming di applicazioni per limitare gli aggiornamenti e i requisiti dei dispositivi.
- Favorisci la collaborazione remota: incoraggia i membri del team a servirsi di strumenti di collaborazione remota come Amazon Chime o AWS Wickr al fine di ridurre la necessità di spostamenti e le emissioni di carbonio associate.
- Usa software a basso consumo energetico: fornisci ai membri del team software a basso consumo energetico, procedendo a rimuovere o disattivare funzionalità e processi non necessari.
- Gestisci i cicli di vita: valuta l'impatto di processi e sistemi sul ciclo di vita dei tuoi dispositivi e seleziona soluzioni che riducono al minimo i requisiti per la sostituzione dei dispositivi, pur

continuando a soddisfare i requisiti di business. Effettua regolarmente la manutenzione e l'aggiornamento delle workstation o del software per conservare e migliorare l'efficienza.

- Gestione remota dei dispositivi: implementa la gestione remota dei dispositivi per ridurre gli spostamenti aziendali.
  - AWS Systems Manager Fleet Manager è un'esperienza di interfaccia utente (UI) unificata che ti aiuta a gestire in remoto i nodi in esecuzione in locale o in locale. AWS

Risorse

Documenti correlati:

- Che cos'è Amazon WorkSpaces?
- Ottimizzatore dei costi per Amazon WorkSpaces
- Documentazione Amazon AppStream 2.0
- NICE DCV

Video correlati:

- Gestione dei costi per Amazon WorkSpaces su AWS

SUS02-BP06 Implementare il buffering o il throttling per appiattire la curva di domanda

Il buffering e la limitazione (della larghezza di banda della rete) riducono la curva delle richieste e la capacità allocata per il tuo carico di lavoro.

Anti-pattern comuni:

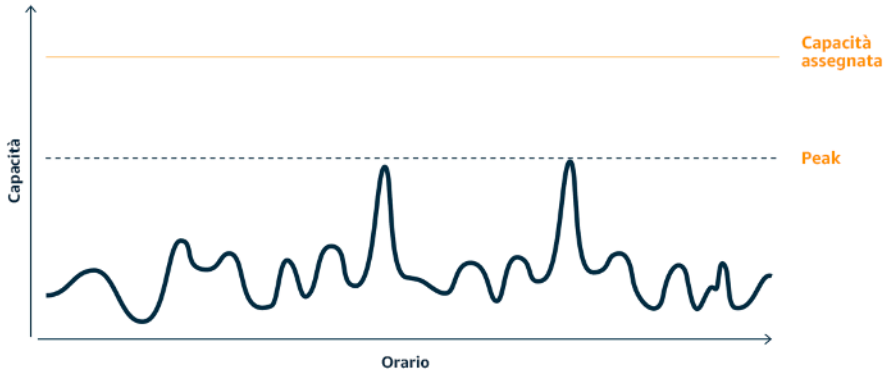
- Elabori immediatamente le richieste del client, anche se non è necessario.
- Non analizzi i requisiti relativi alle richieste dei clienti.

Vantaggi dell'adozione di questa best practice: riduzione della curva della domanda in modo da diminuire la capacità allocata richiesta per il carico di lavoro Ridurre la capacità allocata significa ridurre il consumo di energia e contenere l'impatto ambientale.

Livello di rischio associato se questa best practice non fosse adottata: basso

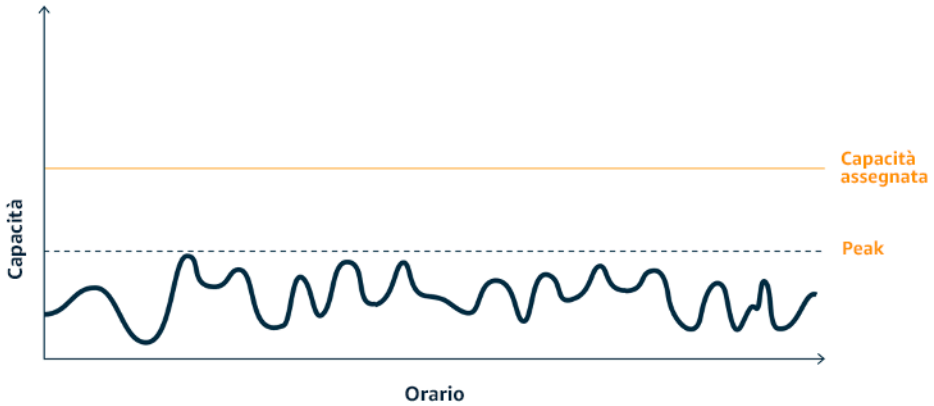
Guida all'implementazione

Diminuire la curva della domanda del carico di lavoro può aiutarti a ridurre la capacità allocata di un carico di lavoro, oltre al suo impatto sull'ambiente. Supponiamo che un carico di lavoro abbia la curva della domanda mostrata nella figura qui sotto. Questo carico di lavoro presenta due picchi e per gestire tali picchi viene eseguito il provisioning della capacità di risorse mostrata dalla linea arancione. Le risorse e l'energia utilizzate per questo carico di lavoro non sono indicate nell'area sotto la curva della domanda, ma nell'area sotto la linea della capacità fornita, poiché per gestire questi due picchi è necessario eseguire il provisioning di tale capacità.



Curva di domanda con due picchi distinti che richiedono un'elevata capacità allocata.

Puoi usare il buffering o la limitazione (della larghezza di banda della rete) per modificare la curva della domanda e appianare i picchi, con conseguente diminuzione della capacità allocata e consumo inferiore di energia. Implementa la limitazione (della larghezza di banda della rete) quando i client eseguono nuovi tentativi. Implementa il buffering per archiviare la richiesta e rinviare l'elaborazione a un secondo momento.



Effetto della limitazione (della larghezza di banda della rete) sulla curva della domanda e sulla capacità allocata.

Passaggi dell'implementazione

- Analizza le richieste del client per stabilire come rispondere. Le domande da considerare includono:
  - Questa richiesta può essere elaborata in modo asincrono?
  - Il client ha la possibilità di ripetere i tentativi?
- Se il client ha la possibilità di ripetere i tentativi puoi implementare la limitazione (della larghezza di banda della rete), che indica alla sorgente che, se non è in grado di soddisfare la richiesta all'ora corrente, dovrebbe riprovare più tardi.
  - Puoi utilizzare Amazon API Gateway per implementare il throttling.
- Per i client che non possono eseguire altri tentativi, è necessario implementare un buffer per ridurre i picchi della curva della domanda. Il buffering rinvia l'elaborazione delle richieste, consentendo alle applicazioni eseguite a velocità diverse di comunicare in modo efficace. Un approccio basato sul buffering impiega una coda o un flusso per l'accettazione dei messaggi dai produttori. I messaggi vengono letti ed elaborati dai consumatori e ciò consente ai messaggi di essere eseguiti alla velocità che soddisfa i requisiti aziendali del consumatore stesso.
  - Amazon Simple Queue Service (AmazonSQS) è un servizio gestito che fornisce code che consentono a un singolo consumatore di leggere singoli messaggi.
  - Amazon Kinesis offre un flusso che consente a più consumatori di leggere gli stessi messaggi.
- Analizza la domanda complessiva, la velocità di modifica e il tempo di risposta richiesto per determinare le dimensioni della limitazione (della larghezza di banda della rete) o del buffer richiesto.

Risorse

Documenti correlati:

- Guida introduttiva ad Amazon SQS
- Application integration Using Queues and Messages
- Gestione e monitoraggio della API limitazione dei carichi di lavoro
- Limitazione su larga scala di un sistema multi-tenant su più livelli utilizzando Gateway REST API API

- Application integration Using Queues and Messages

Video correlati:

- AWS re:Invent 2022 - Modelli di integrazione delle applicazioni per microservizi
- AWS re:Invent 2023 - Risparmio intelligente: strategie di ottimizzazione dei costi di Amazon EC2
- AWS re:Invent 2023 - Modelli di integrazione avanzati e compromessi per sistemi scarsamente accoppiati

Software e architettura

Domanda

- SUS3 Come sfruttate i modelli di software e architettura per supportare i vostri obiettivi di sostenibilità?

SUS3 Come sfruttate i modelli di software e architettura per supportare i vostri obiettivi di sostenibilità?

Implementa modelli per eseguire lo smoothing del carico e garantire un utilizzo elevato e coerente delle risorse implementate per ridurre al minimo il loro consumo. In seguito alle modifiche nei comportamenti degli utenti nel tempo, alcuni componenti potrebbero diventare inattivi per mancanza di utilizzo. Rivedi modelli e architetture per consolidare i componenti sottoutilizzati e aumentare l'uso complessivo. Ritira i componenti non più necessari. Analizza le prestazioni dei componenti dei tuoi carichi di lavoro e ottimizza quelli che usano la maggior quantità di risorse. Identifica i dispositivi che i clienti utilizzano per accedere ai servizi e implementa modelli in grado di ridurre al minimo la necessità di aggiornamenti dei dispositivi.

Best practice

- SUS03-BP01 Ottimizzazione del software e dell'architettura per lavori asincroni e pianificati
- SUS03-BP02 Rimuovere o rifattorizzare i componenti del carico di lavoro con un utilizzo minimo o nullo
- SUS03-BP03 Ottimizza le aree di codice che consumano più tempo o risorse
- SUS03-BP04 Ottimizzazione dell'impatto su dispositivi e apparecchiature
- SUS03-BP05 Utilizza modelli e architetture software che supportano al meglio i modelli di accesso e archiviazione dei dati

SUS03-BP01 Ottimizzazione del software e dell'architettura per lavori asincroni e pianificati

Utilizza modelli efficienti di software e di architettura, come quelli basati sulle code, per mantenere un utilizzo elevato e costante delle risorse distribuite.

Anti-pattern comuni:

- Provisioning di risorse in eccedenza per il carico di lavoro in cloud con lo scopo di far fronte a picchi di domanda imprevisti.
- Architettura non in grado di disaccoppiare i mittenti e i ricevitori di messaggi asincroni mediante un componente di messaggistica.

Vantaggi dell'adozione di questa best practice:

- Modelli efficienti di software e architettura riducono al minimo le risorse inutilizzate nel carico di lavoro e migliorano l'efficienza complessiva.
- È possibile scalare le risorse dedicate all'elaborazione indipendentemente dalla ricezione di messaggi asincroni.
- Grazie a un componente di messaggistica, i requisiti di disponibilità si attenuano e possono essere soddisfatti con un numero inferiore di risorse.

Livello di rischio associato se questa best practice non fosse adottata: medio

Guida all'implementazione

Utilizza modelli di architettura efficienti come l'architettura basata su eventi così da ottenere un utilizzo uniforme dei componenti, oltre alla riduzione al minimo del provisioning eccessivo nel carico di lavoro. L'utilizzo di modelli architetturali efficienti riduce al minimo le risorse inattive a causa del mancato utilizzo dovuto alle variazioni della domanda nel tempo.

Comprendi i requisiti dei componenti del carico di lavoro e adotta modelli di architettura che aumentino l'utilizzo complessivo delle risorse. Ritira i componenti non più necessari.

Passaggi dell'implementazione

- Analizza le esigenze del tuo carico di lavoro per determinare come rispondere a tali richieste.
- Per le richieste o i processi che non necessitano di risposte sincrone, utilizza architetture basate su code e worker a dimensionamento automatico per massimizzare l'utilizzo. Ecco alcuni esempi in cui potresti prendere in considerazione un'architettura basata sulle code:

Meccanismo di accodamento	Descrizione
AWS Batch code di lavoro	AWS Batch i lavori vengono inviati a una coda di lavoro dove risiedono fino a quando non è possibile programmarne l'esecuzione in un ambiente di elaborazione.
Amazon Simple Queue Service e istanze Amazon EC2 Spot	Abbinamento di istanze Amazon SQS e Spot per creare un'architettura efficiente e tollerante ai guasti.

- Per le richieste o i processi che possono essere elaborati in qualsiasi momento, ottieni una maggiore efficienza utilizzando i meccanismi di pianificazione dell'elaborazione delle attività in blocco. Ecco alcuni esempi di meccanismi di pianificazione su: AWS

Meccanismo di pianificazione	Descrizione
Amazon EventBridge Scheduler	Una funzionalità di Amazon EventBridge che ti consente di creare, eseguire e gestire attività pianificate su larga scala.
AWS Glue pianificazione basata sul tempo	Definisci una pianificazione basata sul tempo per i crawler e i lavori in. AWS Glue
Attività pianificate di Amazon Elastic Container Service (AmazonECS)	Amazon ECS supporta la creazione di attività pianificate. Le attività pianificate utilizzano EventBridge le regole di Amazon per eseguire le attività in base a una pianificazione o in risposta a un EventBridge evento.
Instance Scheduler	Configura le pianificazioni di avvio e arresto per le tue istanze di Amazon EC2 e Amazon Relational Database Service.

- Se nella tua architettura utilizzi meccanismi di polling e webhook, sostituiscili con eventi. Utilizza architetture basate sugli eventi per la creazione di carichi di lavoro a elevata efficienza.
- Sfrutta la tecnologia serverless di AWS per eliminare infrastrutture con provisioning eccessivo.

- Dimensiona in modo appropriato i singoli componenti dell'architettura per evitare la presenza di risorse inattive in attesa di input.
  - Puoi sfruttare i suggerimenti per il ridimensionamento corretto in AWS Cost Explorer o AWS Compute Optimizer per individuare le opportunità di dimensionamento corretto.
  - Per ulteriori dettagli, consulta Ridimensionamento corretto: provisioning delle istanze per soddisfare i carichi di lavoro.

Risorse

Documenti correlati:

- What is Amazon Simple Queue Service?
- What is Amazon MQ?
- Scalabilità basata su Amazon SQS
- Che cos'è AWS Step Functions?
- Che cos'è AWS Lambda?
- Utilizzo AWS Lambda con Amazon SQS
- Che cos'è Amazon EventBridge?
- Gestione dei flussi di lavoro asincroni con un REST API

Video correlati:

- AWS re:Invent 2023 - Intraprendiamo il percorso verso un'architettura serverless basata sugli eventi
- AWS re:Invent 2023 - Utilizzo della tecnologia serverless per l'architettura basata sugli eventi e la progettazione basata sul dominio
- AWS re:Invent 2023 - Modelli avanzati basati sugli eventi con Amazon EventBridge
- AWS re:Invent 2023 - Architettura sostenibile: passato, presente e futuro
- Modelli di messaggi asincroni | Eventi AWS

Esempi correlati:

- Architettura basata sugli eventi con processori AWS Graviton e istanze Amazon Spot EC2

SUS03-BP02 Rimuovere o rifattorizzare i componenti del carico di lavoro con un utilizzo minimo o nullo

Elimina i componenti non utilizzati e non più necessari e procedi a rifattorizzare quelli con scarso utilizzo per limitare lo spreco di risorse nel tuo carico di lavoro.

Anti-pattern comuni:

- Non verifichi con regolarità il livello di utilizzo dei singoli componenti del tuo carico di lavoro.
- Non si controllano e analizzano i consigli forniti da strumenti di dimensionamento corretto come. AWS AWS Compute Optimizer

Vantaggi dell'adozione di questa best practice: riduzione al minimo degli sprechi e miglioramento dell'efficienza complessiva del carico di lavoro cloud grazie alla rimozione dei componenti non utilizzati.

Livello di rischio associato se questa best practice non fosse adottata: medio

Guida all'implementazione

Esamina il tuo carico di lavoro per identificare componenti inattivi o non utilizzati. Si tratta di un processo di miglioramento iterativo che può essere attivato da cambiamenti della domanda o dal rilascio di un nuovo servizio cloud. Ad esempio, una riduzione significativa del runtime delle funzioni di AWS Lambda può indicare la necessità di diminuire la dimensione della memoria Inoltre, con il AWS rilascio di nuovi servizi e funzionalità, i servizi e l'architettura ottimali per il carico di lavoro potrebbero cambiare.

Monitora continuamente l'attività del carico di lavoro e cerca le opportunità per migliorare il livello di utilizzo dei singoli componenti. Eliminando i componenti inattivi ed eseguendo attività di ridimensionamento corretto, soddisfi i requisiti aziendali con il numero minimo di risorse cloud.

Passaggi dell'implementazione

- Prepara un inventario delle tue AWS risorse. In AWS, puoi attivarlo Esploratore di risorse AWSper esplorare e organizzare AWS le tue risorse. Per maggiori dettagli, consulta AWS re:Invent 2022 - Come gestire risorse e applicazioni su larga scala. AWS
- Monitora e acquisisci i parametri di utilizzo per i componenti critici del tuo carico di lavoro (come l'utilizzo, CPU l'utilizzo della memoria o il throughput di rete nei parametri Amazon). CloudWatch
- Individua i componenti inutilizzati o sottoutilizzati nell'architettura.

    - Per carichi di lavoro stabili, controlla gli strumenti di AWS dimensionamento corretto, ad esempio a intervalli regolari, per identificare i componenti inattivi, inutilizzati o AWS Compute Optimizersottoutilizzati.
    - Per carichi di lavoro effimeri, valuta metriche di utilizzo per identificare componenti inattivi, inutilizzati o sottoutilizzati.
- Ritira i componenti e gli asset associati (come ECR le immagini di Amazon) che non sono più necessari.
    - Pulizia automatica delle immagini non utilizzate in Amazon ECR
    - Elimina i volumi Amazon Elastic Block Store (AmazonEBS) non utilizzati utilizzando AWS Config e AWS Systems Manager
- Rifattorizza o consolida i componenti sottoutilizzati con altre risorse per promuovere un utilizzo efficiente. Ad esempio, puoi effettuare il provisioning di più piccoli database su una singola istanza di RDS database Amazon invece di eseguire database su singole istanze sottoutilizzate.
- Scopri le risorse allocate dal tuo carico di lavoro per completare un'unità di lavoro.

Risorse

Documenti correlati:

- AWS Trusted Advisor
- Che cos'è Amazon CloudWatch?
- Ridimensionamento corretto: provisioning delle istanze per soddisfare i carichi di lavoro
- Optimizing your cost with Rightsizing Recommendations

Video correlati:

- AWS re:Invent 2023 - Capacità, disponibilità, efficienza in termini di costi: scegline tre

Esempi correlati:

- Ottimizza i modelli hardware e rispetta la sostenibilità KPIs

SUS03-BP03 Ottimizza le aree di codice che consumano più tempo o risorse

Ottimizza il codice eseguito all'interno di diversi componenti della tua architettura per ridurre l'utilizzo delle risorse e massimizzare al tempo stesso le prestazioni.

Anti-pattern comuni:

- Ignori l'ottimizzazione del codice per l'utilizzo delle risorse.
- In genere, rispondi ai problemi di performance aumentando le risorse.
- La revisione del codice e il processo di sviluppo non monitorano le modifiche a livello di performance.

Vantaggi dell'adozione di questa best practice: riduzione al minimo delle risorse utilizzate e ottimizzazione delle prestazioni grazie all'utilizzo di codice efficiente.

Livello di rischio associato se questa best practice non fosse adottata: medio

Guida all'implementazione

È fondamentale esaminare ogni area funzionale, incluso il codice per un'applicazione ideata nel cloud, per ottimizzare l'uso delle risorse e le performance. Monitora costantemente le performance del tuo carico di lavoro negli ambienti di sviluppo e produzione e identifica le opportunità per migliorare gli snippet di codice che comportano un utilizzo particolarmente elevato delle risorse. Adotta un processo di revisione con cadenza regolare per identificare i bug o gli anti-pattern all'interno del codice che utilizzano le risorse in modo non efficiente. Sfrutta algoritmi semplici ed efficienti che hanno gli stessi risultati per il tuo caso d'uso.

Passaggi dell'implementazione

- Utilizza un linguaggio di programmazione efficiente: usa un sistema operativo e un linguaggio di programmazione efficienti per il carico di lavoro. Per dettagli sui linguaggi di programmazione efficienti dal punto di vista delle risorse (incluso Rust), consulta Sustainability with Rust.
- Usa un compagno di codifica AI: prendi in considerazione l'utilizzo di un compagno di codifica AI come Amazon CodeWhisperer per scrivere codice in modo efficiente.
- Automatizza le revisioni del codice: mentre sviluppi i tuoi carichi di lavoro, adotta un processo di revisione del codice automatizzato, per migliorar la qualità e identificare bug e anti-pattern.
  - Automatizza le revisioni del codice con Amazon Reviewer CodeGuru
  - Rilevamento di bug relativi alla concorrenza con Amazon CodeGuru

    - Migliorare la qualità del codice per le applicazioni Python con Amazon CodeGuru
- Usa un profiler di codice: utilizza un profiler di codice per identificare le aree di codice che utilizzano la maggior parte del tempo o delle risorse e trasformale in obiettivi di ottimizzazione.
    - Ridurre l'impronta di carbonio della tua organizzazione con Amazon CodeGuru Profiler
    - Comprendere l'utilizzo della memoria nell'applicazione Java con Amazon CodeGuru Profiler
    - Migliorare l'esperienza dei clienti e ridurre i costi con Amazon CodeGuru Profiler
- Monitora e ottimizza: utilizza risorse di monitoraggio continuo per individuare i componenti con requisiti elevati in termini di risorse o con una configurazione non ottimale.
    - Sostituisci gli algoritmi a uso intensivo di elaborazioni con una versione più semplice ed efficiente che produce gli stessi risultati.
    - Rimuovi il codice non necessario, come quello relativo all'ordinamento e alla formattazione.
- Usa la rifattorizzazione o la trasformazione del codice: scopri le funzionalità di trasformazione del codice Amazon Q per l'esecuzione di manutenzione e aggiornamenti delle applicazioni.
    - Upgrade language versions with Amazon Q Code Transformation
    - AWS re:Invent 2023 - Automatizza gli aggiornamenti e la manutenzione delle app con Amazon Q Code Transformation

Risorse

Documenti correlati:

- Cos'è Amazon CodeGuru Profiler?
- FPGAistanze
- Gli strumenti AWS SDKs su cui costruire AWS

Video correlati:

- Migliora l'efficienza del codice con Amazon CodeGuru Profiler
- AWS re:Invent 2023 - Le migliori pratiche per Amazon CodeWhisperer
- Automatizza le revisioni dei codici e i consigli sulle prestazioni delle applicazioni con Amazon CodeGuru

Esempi correlati:

- Ottimizzazione del codice con Amazon CodeGuru

SUS03-BP04 Ottimizzazione dell'impatto su dispositivi e apparecchiature

Individua i dispositivi e le apparecchiature utilizzati nell'architettura e applica le strategie per ridurne l'utilizzo. Questo può ridurre l'impatto ambientale complessivo del tuo carico di lavoro cloud.

Anti-pattern comuni:

- Ignori l'impatto ambientale dei dispositivi utilizzati dai clienti.
- Gestisci e aggiorni manualmente le risorse utilizzate dai clienti.

Vantaggi della definizione di questa best practice: riduzione dell'impatto ambientale complessivo del carico di lavoro sul cloud grazie all'implementazione di modelli e funzionalità software ottimizzati per i dispositivi dei clienti.

Livello di rischio associato se questa best practice non fosse adottata: medio

Guida all'implementazione

Implementare modelli e funzionalità software ottimizzati per i dispositivi dei clienti può ridurre l'impatto ambientale in diversi modi:

- Implementare nuove funzionalità compatibili con le versioni precedenti può ridurre il numero di sostituzioni hardware.
- Ottimizzare un'applicazione per un'esecuzione ottimale sui dispositivi può contribuire a ridurre l'utilizzo di energia ed estendere la durata della relativa batteria (se alimentati in questo modo).
- Ottimizzare un'applicazione per i dispositivi significa anche ridurre il trasferimento dei dati sulla rete.

Conoscere dispositivi e apparecchiature utilizzati nella tua architettura, il loro ciclo di vita atteso e l'impatto della sostituzione di tali componenti. Implementare modelli e funzionalità software in grado di contribuire a ridurre l'uso di energia da parte del dispositivo, la necessità da parte dei clienti di sostituirlo, nonché di eseguire l'aggiornamento manuale.

Passaggi dell'implementazione

- Predisponi un inventario: fai un inventario dei dispositivi usati nella tua architettura. I dispositivi possono essere mobili, tablet, IOT dispositivi, luci intelligenti o persino dispositivi intelligenti in fabbrica.
- Utilizza dispositivi a basso consumo energetico: prendi in considerazione l'uso dispositivi a basso consumo energetico nella tua architettura. Utilizza le configurazioni di gestione dell'alimentazione sui dispositivi per accedere alla modalità di risparmio energetico quando non sono in uso.
- Esegui applicazioni efficienti: ottimizza l'applicazione in esecuzione sui dispositivi.
  - Usa strategie come l'esecuzione di attività in background per ridurre l'uso di energia.
  - Prendi in considerazione latenza e larghezza di banda della rete durante la creazione di payload e implementa funzionalità che consentano alle tue applicazioni di funzionare in modo ottimale anche in presenza di una larghezza di banda ridotta e di link ad alta latenza.
  - Converti payload e file in formati ottimizzati richiesti dai dispositivi. Ad esempio, puoi usare Amazon Elastic Transcoder o AWS Elemental MediaConvert per convertire file multimediali digitali di alta qualità di grandi dimensioni nei formati utilizzati dagli utenti per la riproduzione su dispositivi mobili, tablet, browser Web e televisioni connesse.
  - Esegui attività a elevata intensità di calcolo lato server (come il rendering delle immagini) oppure usa lo streaming delle applicazioni per migliorare l'esperienza utente sui dispositivi meno recenti.
  - Esegui la segmentazione e la paginazione dell'output, soprattutto per le sessioni interattive, al fine di gestire i payload e limitare i requisiti di archiviazione in locale.
- Coinvolgi i fornitori: collabora con i fornitori dei dispositivi che utilizzano materiali sostenibili e garantiscono trasparenza circa le loro catene di approvvigionamento e certificazioni ambientali.
- Usa over-the-air (OTA) updates: utilizza il meccanismo automatico over-the-air (OTA) per distribuire gli aggiornamenti su uno o più dispositivi.
  - Per aggiornare le applicazioni mobili, puoi utilizzare una pipeline CI/CD.
  - Puoi usare AWS IoT Device Management per gestire in remoto i dispositivi connessi su larga scala.
- Usa device farm gestite: per testare nuove funzionalità e aggiornamenti, usa device farm gestite con set di hardware rappresentativi e itera lo sviluppo per ottimizzare i dispositivi supportati. Per ulteriori dettagli, consulta SUS06-BP04 Usa farm di dispositivi gestiti per i test.
- Continua a monitorare e apportare miglioramenti: monitora il consumo energetico dei dispositivi per identificare le aree di miglioramento. Utilizza le nuove tecnologie o best practice per migliorare l'impatto ambientale di tali dispositivi.

Risorse

Documenti correlati:

- Che cos'è AWS Device Farm?
- AppStream Documentazione 2.0
- NICE DCV
- OTAtutorial per l'aggiornamento del firmware sui dispositivi che eseguono Free RTOS
- Optimizing Your IoT Devices for Environmental Sustainability

Video correlati:

- AWS re:Invent 2023 - Migliora la qualità delle tue app per dispositivi mobili e web utilizzando AWS Device Farm

SUS03-BP05 Utilizza modelli e architetture software che supportano al meglio i modelli di accesso e archiviazione dei dati

Scopri come i dati vengono utilizzati all'interno del tuo carico di lavoro, consumati dagli utenti, trasferiti e archiviati. Usa architetture e modelli software in grado di supportare al meglio l'accesso ai dati e l'archiviazione per ridurre le risorse di elaborazione, rete e storage richieste dal carico di lavoro.

Anti-pattern comuni:

- Ritieni che tutti i carichi di lavoro abbiano modelli di accesso e archiviazione di dati simili.
- Utilizzi un solo livello di archiviazione, presupponendo che tutti i carichi di lavoro rientrino in tale livello.
- Ritieni che gli schemi di accesso ai dati rimarranno coerenti nel tempo.
- La tua architettura supporta una potenziale espansione elevata dell'accesso ai dati, con conseguente inattività delle risorse per la maggior parte del tempo.

Vantaggi dell'adozione di questa best practice: riduzione della complessità dello sviluppo e aumento dell'utilizzo complessivo grazie alla selezione e all'ottimizzazione dell'architettura in base ai modelli di accesso ai dati e di archiviazione. Capire quando utilizzare le tabelle globali, il partizionamento dei dati e la memorizzazione nella cache, ti aiuterà a ridurre i costi operativi e a effettuare il dimensionamento in base alle esigenze del carico di lavoro.

Livello di rischio associato se questa best practice non fosse adottata: medio

Guida all'implementazione

Usa modelli di software e architetture che siano quanto più in linea con le caratteristiche dei tuoi dati e i modelli di accesso. Ad esempio, utilizza un'architettura dati moderna su AWS in modo da utilizzare servizi appositamente progettati e ottimizzati per i tuoi casi d'uso di analisi unici. Questi modelli di architettura consentono un'elaborazione efficiente dei dati e riducono l'utilizzo delle risorse.

Passaggi dell'implementazione

- Analizza le caratteristiche dei dati e i modelli di accesso per individuare la configurazione corretta per le tue risorse cloud. Gli aspetti chiave da considerare includono:
  - Tipo di dati: strutturati, semi-strutturati, non strutturati
  - Crescita dei dati: limitata, illimitata
  - Durabilità dei dati: persistenti, effimeri, transitori
  - Schemi di accesso: letture o scritture, frequenza di aggiornamento, con picchi o costante
- Usa tipi di architetture che meglio supportano l'accesso ai dati e i modelli di archiviazione.
  - Patterns for enabling data persistence
  - Let's Architect! Modern data architectures
  - Database su AWS: Lo strumento giusto per il giusto Job
- Sfrutta le tecnologie che lavorano in modo nativo con i dati compressi.
  - Athena Compression Support file formats
  - Opzioni di formato per ETL ingressi e uscite in AWS Glue
  - Loading compressed data files from Amazon S3 with Amazon Redshift
- Sfrutta servizi di analisi appositamente creati per l'elaborazione dei dati nella tua architettura. Per informazioni dettagliate sui servizi di analisi AWS appositamente progettati, consulta AWS re:Invent 2022 - Building modern data architectures on. AWS
- Utilizza il motore del database che meglio supporta il modello di query dominante. Gestisci gli indici di database per garantire un'esecuzione efficiente delle query. Per ulteriori informazioni, consulta Database su AWS e guarda AWS re:Invent 2022 - Modernize apps with purpose-built databases.
- Seleziona protocolli di rete che riducano la quantità di capacità di rete utilizzata dalla tua architettura.

Risorse

Documenti correlati:

- COPYda formati di dati colonnari con Amazon Redshift
- Converting Your Input Record Format in Firehose
- Migliora le prestazioni delle query su Amazon Athena con una conversione ai formati in colonne
- Monitoring DB load with Performance Insights on Amazon Aurora
- Monitoraggio del carico del DB con Performance Insights su Amazon RDS
- Classe di archiviazione del Piano intelligente Amazon S3
- Crea un negozio di CQRS eventi con Amazon DynamoDB

Video correlati:

- AWS re:Invent 2022 - Creazione di architetture di data mesh su AWS
- AWS re:Invent 2023 - Approfondisci Amazon Aurora e le sue innovazioni
- AWS re:Invent 2023 - Migliora l'EBSefficienza di Amazon e sii più efficiente in termini di costi
- AWS re:Invent 2023 - Ottimizzazione del prezzo e delle prestazioni dello storage con Amazon S3
- AWS re:Invent 2023 - Creazione e ottimizzazione di un data lake su Amazon S3
- AWS re:Invent 2023 - Modelli avanzati basati sugli eventi con Amazon EventBridge

Esempi correlati:

- AWS Workshop sui database appositamente progettati
- AWS Giornata di immersione nell'architettura dei dati moderna
- Crea una rete di dati su AWS

Dati

Domanda

- SUS4 Come sfruttate le politiche e i modelli di gestione dei dati per supportare i vostri obiettivi di sostenibilità?

SUS4 Come sfruttate le politiche e i modelli di gestione dei dati per supportare i vostri obiettivi di sostenibilità?

Implementa procedure di gestione dei dati per ridurre l'archiviazione allocata richiesta per supportare il carico di lavoro e le risorse necessarie per l'uso correlato. Analizza i tuoi dati e usa tecnologie e configurazioni di archiviazione che supportano più efficacemente il valore aziendale dei dati e il modo in cui vengono utilizzati. Esegui il ciclo di vita dei dati su un'archiviazione più efficiente e meno performante al diminuire dei requisiti ed elimina i dati che non sono più necessari.

Best practice

- SUS04-BP01 Implementare una politica di classificazione dei dati
- SUS04-BP02 Utilizza tecnologie che supportano l'accesso ai dati e i modelli di archiviazione
- SUS04-BP03 Usa le policy per gestire il ciclo di vita dei tuoi set di dati
- SUS04-BP04 Usa l'elasticità e l'automazione per espandere lo storage a blocchi o il file system
- SUS04-BP05 Rimuovere i dati non necessari o ridondanti
- SUS04-BP06 Usa file system o storage condivisi per accedere a dati comuni
- SUS04-BP07 Ridurre al minimo lo spostamento dei dati tra le reti
- SUS04-BP08 Esegui il backup dei dati solo quando è difficile ricrearli

SUS04-BP01 Implementare una politica di classificazione dei dati

Classifica i dati per capire le criticità rispetto ai risultati aziendali e scegli il livello di archiviazione ad alta efficienza corretto per le tue informazioni.

Anti-pattern comuni:

- Non identifichi asset di dati con caratteristiche simili (come sensibilità, criticità aziendale o requisiti normativi) che vengono elaborati o archiviati.
- Non hai implementato un catalogo di dati per eseguire l'inventario dei tuoi asset.

Vantaggi dell'adozione di questa best practice: determinazione del livello di archiviazione dei dati più efficiente dal punto di vista energetico grazie all'implementazione di una policy di classificazione dei dati.

Livello di rischio associato se questa best practice non fosse adottata: medio

Guida all'implementazione

La classificazione dei dati comporta l'identificazione dei tipi di dati elaborati e archiviati in un sistema informativo di proprietà o gestito da un'organizzazione. Inoltre, è necessario stabilire la criticità dei dati e il probabile impatto di una compromissione, perdita o uso improprio dei dati.

Implementare la policy di classificazione dei dati partendo dall'uso contestuale dei dati e creando uno schema di categorizzazione che tenga conto del livello di criticità di un determinato set di dati per le operazioni dell'organizzazione.

Passaggi dell'implementazione

- Esegui l'inventario dei dati: redigi l'inventario dei vari tipi di dati esistenti per il carico di lavoro.
- Raggruppa i dati: determina la criticità, la riservatezza, l'integrità e la disponibilità dei dati in base al rischio per l'organizzazione. Utilizza questi requisiti per raggruppare i dati in uno dei livelli di classificazione dei dati adottati. Ad esempio, consulta Quattro semplici passaggi per classificare i dati e proteggere la tua startup.
- Definisci livelli di classificazione dei dati e policy: per ciascun gruppo di dati, definisci il livello di classificazione dei dati (ad esempio, pubblico o riservato) e le policy di gestione. Applica ai dati i tag adeguati. Per maggiori dettagli sulle categorie di classificazione dei dati, consulta il whitepaper sulla classificazione dei dati.
- Rivedi periodicamente: esamina e controlla periodicamente l'ambiente per verificare la presenza di dati senza tag e non classificati. Usa l'automazione per identificare questi dati, classificandoli e applicando i tag in modo appropriato. Ad esempio, consulta Data Catalog and crawlers in AWS Glue.
- Crea un catalogo dati: definisci un catalogo dati con funzionalità di audit e governance
- Documenta: crea documenti relativi a policy di classificazione dei dati e procedure di gestione per ciascuna classe di dati.

Risorse

Documenti correlati:

- Leveraging Cloud AWS to Support Data Classification
- Politiche di tag da AWS Organizations

Video correlati:

- AWS re:Invent 2022 - Promuovere l'agilità con la governance dei dati attiva AWS
- AWS re:Invent 2023 - Protezione e resilienza dei dati con storage AWS

SUS04-BP02 Utilizza tecnologie che supportano l'accesso ai dati e i modelli di archiviazione

Usa tecnologie di archiviazione in grado di supportare al meglio il modo in cui viene effettuato l'accesso ai dati e come vengono archiviati per ridurre la quantità di risorse allocate e supportare al tempo stesso il tuo carico di lavoro.

Anti-pattern comuni:

- Ritieni che tutti i carichi di lavoro abbiano modelli di accesso e archiviazione di dati simili.
- Utilizzi un solo livello di archiviazione, presupponendo che tutti i carichi di lavoro rientrino in tale livello.
- Ritieni che gli schemi di accesso ai dati rimarranno coerenti nel tempo.

Vantaggi dell'adozione di questa best practice: selezionare e ottimizzare le tecnologie di archiviazione in base all'accesso ai dati e ai modelli di archiviazione ti consentirà di ridurre le risorse cloud richieste per soddisfare le tue esigenze aziendali e migliorare l'efficienza generale del tuo carico di lavoro cloud.

Livello di rischio associato se questa best practice non fosse adottata: basso

Guida all'implementazione

Seleziona la soluzione di archiviazione più adatta ai tuoi modelli di accesso. In alternativa, puoi modificarli affinché siano in linea con la soluzione di archiviazione, allo scopo di ottimizzare l'efficienza delle prestazioni.

Passaggi dell'implementazione

- Esamina le caratteristiche dei dati e dell'accesso: valuta le caratteristiche dei tuoi dati e il modello di accesso per raccogliere le caratteristiche chiave delle tue esigenze di archiviazione. Gli aspetti chiave da considerare includono:
  - Tipo di dati: strutturati, semi-strutturati, non strutturati
  - Crescita dei dati: limitata, illimitata
  - Durabilità dei dati: persistenti, effimeri, transitori

  - Modelli di accesso: letture o scritture, frequenza, con picchi o costante
- Scegli la giusta tecnologia di archiviazione: migra i dati alla tecnologia di archiviazione appropriata che supporta le caratteristiche dei tuoi dati e il modello di accesso. Ecco alcuni esempi di tecnologie di AWS storage e le relative caratteristiche principali:

Tipo	Tecnologia	Caratteristiche chiave
Archiviazione di oggetti	Amazon S3	Un servizio di archiviazione di oggetti con scalabilità illimitata, elevata disponibilità e più opzioni di accessibilità. Il trasferimento di oggetti e il relativo trasferimento da e verso Amazon S3 può utilizzare un servizio, come Transfer Acceleration o Punti di accesso, per supportare la posizione, le esigenze di sicurezza e i modelli di accesso.
Archiviazione	Amazon S3 Glacier	Classe di archiviazione di Amazon S3 creata per l'archiviazione dei dati.
File system condiviso	Amazon Elastic File System (AmazonEFS)	File system montabile a cui è possibile accedere da più tipi di soluzioni di calcolo. Amazon aumenta e riduce EFS automaticamente lo storage ed è ottimizzato per le prestazioni per offrire latenze basse e costanti.
File system condiviso	Amazon FSx	Basato sulle più recenti soluzioni di AWS elaborazione per supportare quattro

Tipo	Tecnologia	Caratteristiche chiave
		file system di uso comune: Open NetApp ONTAPZFS, Windows File Server e Lustre. FSxLa latenza, la velocità effettiva e la velocità effettiva di Amazon IOPS variano in base al file system e devono essere prese in considera zione quando si seleziona il file system giusto per le esigenze di carico di lavoro.
Storage a blocchi	Amazon Elastic Block Store (AmazonEBS)	Servizio di storage a blocchi scalabile e ad alte prestazio ni progettato per Amazon Elastic Compute Cloud (Amazon). EC2 Amazon EBS include storage SSD supportato per carichi di lavoro transazionali e intensivi e HDD storage supportat o per carichi di lavoro con throughput IOPS intensivo.

Tipo	Tecnologia	Caratteristiche chiave
Database relazionale	Amazon Aurora, AmazonRDS, Amazon Redshift	Progettato per supportare transazioni ACID (atomicità, coerenza, isolamento, durabilità) e mantenere l'integrità referenziale e una forte coerenza dei dati. Molte applicazioni tradizionali, i sistemi di pianificazione delle risorse aziendali (ERP), di gestione delle relazioni con i clienti (CRM) e di e-commerce utilizzano database relazionali per archiviare i propri dati.
Database chiave-valore	Amazon DynamoDB	Ottimizzato per schemi di accesso di uso comune, in genere per archiviare e recuperare grandi volumi di dati. Le app Web dal traffico elevato, i sistemi di e-commerce e le applicazioni di videogiochi sono casi d'uso tipici dei database chiave-valore.

- Automatizza l'allocazione dello storage: per i sistemi di storage di dimensioni fisse, come Amazon o EBS AmazonFSx, monitora lo spazio di archiviazione disponibile e automatizza l'allocazione dello storage al raggiungimento di una soglia. Puoi sfruttare Amazon CloudWatch per raccogliere e analizzare diversi parametri per Amazon e EBSAmazon. FSx
- Scegli la classe di archiviazione giusta: scegli la classe di archiviazione opportuna per i tuoi dati.
  - Le classi di archiviazione Amazon S3 possono essere configurate a livello di oggetto. Un singolo bucket può contenere oggetti archiviati per tutte le classi di archiviazione.
  - Puoi utilizzare le policy del ciclo di vita Amazon S3 per passare automaticamente gli oggetti tra le classi di archiviazione oppure rimuovere i dati senza modifiche all'applicazione. In generale,

devi raggiungere un equilibrio tra efficienza delle risorse, latenza di accesso e affidabilità, quando consideri questi meccanismi di storage.

Risorse

Documenti correlati:

- Tipi di EBS volume Amazon
- Amazon EC2 Instance Store
- Amazon S3 Intelligent-Tiering
- Caratteristiche di Amazon EBS I/O
- Using Amazon S3 storage classes
- What is Amazon S3 Glacier?

Video correlati:

- AWS re:Invent 2023 - Migliora l'EBSefficienza di Amazon e sii più efficiente in termini di costi
- AWS re:Invent 2023 - Ottimizzazione del prezzo e delle prestazioni dello storage con Amazon S3
- AWS re:Invent 2023 - Creazione e ottimizzazione di un data lake su Amazon S3
- AWS re:Invent 2022 - Creazione di moderne architetture di dati su AWS
- AWS re:Invent 2022 - Modernizza le app con database creati appositamente
- AWS re:Invent 2022 - Creazione di architetture di data mesh su AWS
- AWS re:Invent 2023 - Approfondisci Amazon Aurora e le sue innovazioni
- AWS re:Invent 2023 - Modellazione avanzata dei dati con Amazon DynamoDB

Esempi correlati:

- Amazon S3 Examples
- AWS Workshop su database appositamente progettati
- Databases for Developers
- AWS Giornata di immersione nell'architettura dei dati moderna
- Crea una rete di dati su AWS

SUS04-BP03 Usa le policy per gestire il ciclo di vita dei tuoi set di dati

Gestisci il ciclo di vita di tutti i tuoi dati e applica in automatico le cancellazioni per ridurre i requisiti totali di archiviazione del tuo carico di lavoro.

Anti-pattern comuni:

- Cancellazione manuale dei dati.
- Conservazione di tutti i dati del carico di lavoro.
- Mancato spostamento dei dati su livelli di archiviazione più efficienti dal punto di vista energetico in base ai requisiti di conservazione e accesso.

Vantaggi dell'adozione di questa best practice: l'utilizzo delle policy per il ciclo di vita dei dati garantisce un accesso e una conservazione efficienti dei dati in un carico di lavoro.

Livello di rischio associato se questa best practice non fosse adottata: medio

Guida all'implementazione

i set di dati presentano solitamente requisiti di conservazione e accesso che cambiano durante il loro ciclo di vita. Ad esempio, l'applicazione potrebbe avere bisogno di accedere frequentemente ad alcuni set di dati per un periodo di tempo limitato. In seguito, questi set di dati vengono consultati di rado.

Per gestire in modo efficiente i set di dati durante il loro ciclo di vita, è necessario configurare le policy per il ciclo di vita, ovvero le regole che definiscono la gestione dei set di dati.

Con le regole di configurazione del ciclo di vita, è possibile indicare al servizio di archiviazione di trasferire un set di dati a livelli di archiviazione più efficienti dal punto di vista energetico, di archiviarlo o di eliminarlo.

Passaggi dell'implementazione

- Classifica i set di dati del carico di lavoro.
- Definisci le procedure di gestione per ogni classe di dati.
- Imposta policy automatizzate per il ciclo di vita per applicare le regole correlate. Ecco alcuni esempi di come configurare politiche automatizzate del ciclo di vita per diversi servizi di storage: AWS

Servizio di storage	Come impostare policy automatizzate per il ciclo di vita
Amazon S3	Puoi utilizzare il ciclo di vita Amazon S3 per gestire gli oggetti durante il loro ciclo di vita. In caso di schemi di accesso sconosciuti, mutevoli o imprevedibili, puoi utilizzare il Piano intelligente Amazon S3, che monitora gli schemi di accesso e sposta in automatico gli oggetti che non hanno fatto registrare accessi a livelli di accessi più economici. Sfrutta i parametri di Amazon S3 Storage Lens per individuare opportunità di ottimizzazione e lacune nella gestione del ciclo di vita.
Amazon Elastic Block Store	Puoi utilizzare Amazon Data Lifecycle Manager per automatizzare la creazione, la conservazione e l'eliminazione di istantanee EBS Amazon e Amazon -backed. EBS AMIs
Amazon Elastic File System	Amazon EFS Lifecycle Management gestisce automaticamente lo storage dei file per i tuoi file system.
Amazon Elastic Container Registry	Le politiche ECR del ciclo di vita di Amazon automatizzano la pulizia delle immagini dei container facendo scadere le immagini in base all'età o al numero.
AWS Elemental MediaStore	Puoi utilizzare una politica del ciclo di vita degli oggetti che regola per quanto tempo gli oggetti devono essere conservati nel contenito re. MediaStore

- Elimina i volumi inutilizzati, gli snapshot e i dati che hanno superato il periodo di conservazione. Sfrutta le funzionalità di servizio native come Amazon DynamoDB Time To Live o la conservazione dei log di CloudWatch Amazon per l'eliminazione.

- Aggrega e comprimi i dati quando possibile in base alle regole del ciclo di vita.

Risorse

Documenti correlati:

- Ottimizzazione delle regole del ciclo di vita di Amazon S3 con Amazon S3 Storage Class Analysis
- Valutazione delle risorse con Regole di AWS Config

Video correlati:

- AWS re:Invent 2021 - Le migliori pratiche del ciclo di vita di Amazon S3 per ottimizzare la spesa di storage
- AWS re:Invent 2023 - Ottimizzazione del prezzo e delle prestazioni dello storage con Amazon S3
- Simplify Your Data Lifecycle and Optimize Storage Costs With Amazon S3 Lifecycle
- Reduce Your Storage Costs Using Amazon S3 Storage Lens

SUS04-BP04 Usa l'elasticità e l'automazione per espandere lo storage a blocchi o il file system

Usa l'elasticità e l'automazione per espandere lo storage a blocchi o il file system con l'aumento dei dati per ridurre l'archiviazione allocata.

Anti-pattern comuni:

- Acquisti uno storage a blocchi di grandi dimensioni o un file system per necessità future.
- Il numero di operazioni di input e output al secondo (IOPS) del file system è superiore al numero di operazioni di input e output.
- Non monitori l'utilizzo dei volumi di dati.

Vantaggi dell'adozione di questa best practice: la riduzione del provisioning eccessivo per il sistema di archiviazione riduce le risorse inattive e migliora l'efficienza complessiva del tuo carico di lavoro.

Livello di rischio associato se questa best practice non fosse adottata: medio

Guida all'implementazione

Crea storage a blocchi e file system con l'allocazione delle dimensioni, il throughput e la latenza adeguati al tuo carico di lavoro. Usa l'elasticità e l'automazione per espandere lo storage a blocchi

o il file system con l'aumento dei dati per evitare un provisioning eccessivo per questi servizi di archiviazione.

Passaggi dell'implementazione

- Per lo storage a dimensione fissa come Amazon EBS, verifica di monitorare la quantità di storage utilizzata rispetto alla dimensione complessiva dello storage e crea l'automazione, se possibile, per aumentare le dimensioni dello storage quando si raggiunge una soglia.
- Utilizza volumi elastici e servizi di dati a blocchi gestiti per automatizzare l'allocazione di archivi aggiuntivi man mano che i dati persistenti aumentano. Ad esempio, puoi utilizzare Amazon EBS Elastic Volumes per modificare la dimensione del volume, il tipo di volume o regolare le prestazioni dei tuoi EBS volumi Amazon.
- Scegli la classe di archiviazione corretta, le performance e il throughput per il tuo file system per rispondere alle esigenze della tua azienda, senza eccedere.
  - EFSPrestazioni di Amazon
  - Prestazioni dei EBS volumi Amazon su istanze Linux
- Imposta i livelli target di utilizzo per i volumi di dati e ridimensiona i volumi al di fuori degli intervalli previsti.
- Dimensiona i volumi di sola lettura per adattarli ai dati.
- Migra i dati su archivi oggetti per evitare il provisioning di capacità eccessive da dimensioni di volumi fisse su archiviazioni a blocchi.
- Esamina regolarmente i volumi elastici e i file system per terminare i volumi inattivi e ridurre i volumi con un provisioning eccessivo per adattarli alla dimensione corrente dei dati.

Risorse

Documenti correlati:

- Estendi il file system dopo il ridimensionamento di un volume EBS
- Modifica un volume utilizzando Amazon EBS Elastic Volumes
- Documentazione FSx Amazon
- What is Amazon Elastic File System?

Video correlati:

- Approfondimento su Amazon EBS Elastic Volumes
- Strategie di ottimizzazione di Amazon EBS e Snapshot per migliori prestazioni e risparmi sui costi
- Ottimizzazione di Amazon in termini EFS di costi e prestazioni, utilizzando le best practice

SUS04-BP05 Rimuovere i dati non necessari o ridondanti

Elimina i dati non necessari o ridondanti per ridurre al minimo le risorse di archiviazione necessarie per memorizzare i set di dati.

Anti-pattern comuni:

- Duplicazione dei dati che possono essere facilmente recuperati o ricreati.
- Backup di tutti i dati senza prenderne in considerazione la criticità.
- Cancellazione dei dati eseguita in modo irregolare, in occasione di eventi operativi o non eseguita affatto.
- Archiviazione dei dati in modo ridondante, indipendentemente dall'affidabilità del servizio di archiviazione.
- Attivazione del controllo delle versioni di Amazon S3 senza alcuna giustificazione aziendale.

Vantaggi dell'adozione di questa best practice: riduzione delle dimensioni di archiviazione necessarie per il carico di lavoro e del suo impatto ambientale grazie alla rimozione dei dati non necessari.

Livello di rischio associato se questa best practice non fosse adottata: medio

Guida all'implementazione

Non memorizzare i dati che non ti servono. Automatizza l'eliminazione dei dati non necessari. Utilizza tecnologie di backup che deduplicano i dati a livello di file e blocco. Sfrutta le funzionalità native di replica e ridondanza dei dati dei servizi.

Passaggi dell'implementazione

- Valuta la possibilità di non archiviare i dati utilizzando i set di dati esistenti disponibili al pubblico in AWS Data Exchange e Open Data su AWS.
- Utilizza meccanismi che possano deduplicare i dati a livello di blocco e oggetto. Ecco alcuni esempi di come deduplicare i dati su: AWS

Servizio di storage	Meccanismi di deduplicazione
Amazon S3	AWS Lake Formation FindMatchesUtilizzatelo per trovare i record corrispondenti in un set di dati (compresi quelli senza identificatori) utilizzando il nuovo ML Transform. FindMatches
Amazon FSx	Usa la deduplicazione dei dati su Amazon FSx per Windows.
Volumi e snapshot di Amazon Elastic Block Store	Gli snapshot sono incrementali, ovvero vengono salvati solo i blocchi sul dispositivo che sono cambiati dall'ultimo snapshot.

- Analizza l'accesso ai dati per identificare quelli non necessari. Automatizza le policy per il ciclo di vita. Sfrutta funzionalità di servizio native come Amazon DynamoDB Time To Live, AmazonS3 Lifecycle o Amazon Log Retention per l'eliminazione. CloudWatch
- Utilizza le funzionalità di virtualizzazione dei dati AWS per mantenere i dati alla fonte ed evitare la duplicazione dei dati.
  - Virtualizzazione dei dati nativa per il cloud su AWS
  - Optimize Data Pattern Using Amazon Redshift Data Sharing
- Utilizza una tecnologia di backup in grado di eseguire backup incrementali.
- Sfrutta la durabilità di Amazon S3 e la replica di EBS Amazon per raggiungere i tuoi obiettivi di durabilità invece delle tecnologie autogestite (come una serie ridondante di dischi indipendenti ()). RAID
- Centralizza i log e traccia i dati, deduplica le voci di log identiche e stabilisci meccanismi per ottimizzarne la verbosità quando necessario.
- Popola in anticipo le cache solo quando è necessario.
- Definisci il monitoraggio e l'automazione della cache per ridimensionarla in base alle esigenze.
- Rimuovi le out-of-date distribuzioni e le risorse dagli archivi di oggetti e dalle cache edge quando inserisci nuove versioni del tuo carico di lavoro.

Risorse

Documenti correlati:

- Modifica la conservazione dei dati di registro in Logs CloudWatch
- Deduplicazione dei dati su Amazon FSx per Windows File Server
- Funzionalità di Amazon FSx per l'ONTAPinclusione della deduplicazione dei dati
- Invalidazione dei file su Amazon CloudFront
- Utilizzo AWS Backup per il backup e il ripristino dei EFS file system Amazon
- Che cos'è Amazon CloudWatch Logs?
- Lavorare con i backup su Amazon RDS
- Integra e deduplica i set di dati utilizzando AWS Lake Formation

Video correlati:

- Amazon Redshift Data Sharing Use Cases

Esempi correlati:

- Come posso usare Amazon Athena per analizzare i log di accesso al server Amazon S3?

SUS04-BP06 Usa file system o storage condivisi per accedere a dati comuni

Adotta file system condivisi o l'archiviazione per evitare duplicazioni di dati e abilitare un'infrastruttura più efficiente per il tuo carico di lavoro.

Anti-pattern comuni:

- Esegui il provisioning dell'archiviazione per ogni singolo client.
- Non scolleghi volumi di dati da client inattivi.
- Non fornisci l'accesso allo storage su piattaforme e sistemi.

Vantaggi dell'adozione di questa best practice: condivisione di dati con uno o più utenti senza la necessità di copiarli grazie all'utilizzo di file system o archiviazione condivisi. Questo consente di ridurre le risorse di archiviazione necessarie per il carico di lavoro.

Livello di rischio associato se questa best practice non fosse adottata: medio

Guida all'implementazione

Se hai più utenti o applicazioni che accedono agli stessi set di dati, usare una tecnologia di archiviazione condivisa è fondamentale per usare un'infrastruttura efficiente per il tuo carico di lavoro. La tecnologia di archiviazione condivisa offre una posizione centrale per archiviare e gestire set di dati ed evitare la loro duplicazione. Verifica anche la coerenza dei dati su sistemi diversi. Inoltre, la tecnologia di archiviazione condivisa consente un uso più efficiente della potenza di elaborazione, poiché più risorse di calcolo possono accedere ed elaborare i dati allo stesso momento in parallelo.

Acquisisci i dati dai servizi di archiviazione condivisa in base alle necessità e scollega i volumi non utilizzati per liberare le risorse.

Passaggi dell'implementazione

- Esegui la migrazione dei dati nell'archiviazione condivisa quando i dati hanno più consumer. Ecco alcuni esempi di tecnologia di storage condiviso su: AWS

Opzione di archiviazione	Quando usare
Amazon con collegamento EBS multiplo	Amazon EBS Multi-Attach consente di collegare un singolo volume Provisioned IOPS SSD (io1 o io2) a più istanze che si trovano nella stessa zona di disponibilità.
Amazon EFS	Scopri Quando scegliere Amazon EFS.
Amazon FSx	Vedi Scelta di un FSx file system Amazon.
Amazon S3	Le applicazioni che non richiedono una struttura di file system e sono progettate per lavorare con lo storage degli oggetti possono usare Amazon S3 come soluzione di archiviaz ione degli oggetti a basso costo, durevole e altamente scalabile.

- Copia o acquisisci i dati solo da file system condivisi in base alle necessità. Ad esempio, puoi creare un file system Amazon FSx for Lustre supportato da Amazon S3 e caricare su Amazon solo il sottoinsieme di dati necessario per i processi di elaborazione. FSx

- Elimina i dati nella modalità corretta per i tuoi modelli di utilizzo come illustrato in SUS04-BP03 Usa le policy per gestire il ciclo di vita dei tuoi set di dati.
- Distacca i volumi dai client che non li utilizzano attivamente.

Risorse

Documenti correlati:

- Linking your file system to an Amazon S3 bucket
- Utilizzo di Amazon EFS per AWS Lambda le tue applicazioni serverless
- Amazon EFS Intelligent-Tiering ottimizza i costi per i carichi di lavoro modificando i modelli di accesso
- Utilizzo di Amazon FSx con il tuo repository di dati locale

video correlati:

- Ottimizzazione dei costi di storage con Amazon EFS
- AWS re:Invent 2023 - Cosa c'è di nuovo con lo storage di file AWS
- AWS re:Invent 2023 - Archiviazione di file per costruttori e data scientist su Amazon Elastic File System

SUS04-BP07 Ridurre al minimo lo spostamento dei dati tra le reti

Usa file system condivisi o lo storage a oggetti per accedere ai dati comuni e contenere le risorse di rete totali necessarie per supportare i trasferimenti dei dati per il carico di lavoro.

Anti-pattern comuni:

- Tutti i dati vengono archiviati nello stesso spazio Regione AWS indipendentemente da dove si trovano gli utenti dei dati.
- Non ottimizzi la dimensione e il formato dei dati prima di trasferirli sulla rete.

Vantaggi dell'adozione di questa best practice: l'ottimizzazione del trasferimento dei dati sulla rete riduce la quantità di risorse di rete totali richieste per il carico di lavoro e diminuisce l'impatto ambientale.

Livello di rischio associato se questa best practice non fosse adottata: medio

Guida all'implementazione

Trasferire i dati all'interno dell'organizzazione significa disporre di risorse di elaborazione, rete e archiviazione. Usa tecniche per ridurre il movimento dei dati e migliorare l'efficienza generale del tuo carico di lavoro.

Passaggi dell'implementazione

- Considera la vicinanza ai dati o agli utenti come fattore decisivo per la seleziona di una regione per il tuo carico di lavoro.
- Esegui la partizione dei servizi consumati a livello regionale in modo che i dati specifici della regione siano archiviati nella regione in cui sono usati.
- Utilizza formati di file efficienti (come Parquet oORC) e comprimi i dati prima di spostarli sulla rete.
- Non trasferire dati inutilizzati. Alcuni esempi che possono aiutarti a evitare di spostare dati inutilizzati:
  - Riduci API le risposte ai soli dati pertinenti.
  - Aggrega i dati laddove richiesto (le informazioni a livello di record non sono necessarie).
  - Consulta Well-Architected Lab - Optimize Data Pattern Using Amazon Redshift Data Sharing.
  - Prendi in considerazione la condivisione dei dati tra account in AWS Lake Formation.
- Utilizza servizi in grado di supportarti nell'esecuzione del codice in posizioni più vicine agli utenti del carico di lavoro.

Servizio	Quando usare
Lambda@Edge	Da utilizzare per operazioni a uso intensivo di risorse di calcolo eseguite quando gli oggetti non si trovano nella cache.
CloudFrontFunzioni	Da utilizzare per casi d'uso semplici come manipolazioni di HTTP richiesta/risposta che possono essere avviate da funzioni di breve durata.
AWS IoT Greengrass	Eseguire la memorizzazione nella cache di risorse di calcolo, messaggistica e dati per i dispositivi connessi.

Risorse

Documenti correlati:

- Ottimizzazione dell' AWS infrastruttura per la sostenibilità, parte: rete III
- AWS Infrastruttura globale
- Caratteristiche CloudFront principali di Amazon, tra cui la rete CloudFront Global Edge
- Compressione HTTP delle richieste in Amazon OpenSearch Service
- Compressione intermedia dei dati con Amazon EMR
- Caricamento di file di dati compressi da Amazon S3 a Amazon Redshift
- Servire file compressi con Amazon CloudFront

Video correlati:

- Demistificare il trasferimento di dati su AWS

Esempi correlati:

- Architecting for sustainability - Minimize data movement across networks

SUS04-BP08 Esegui il backup dei dati solo quando è difficile ricrearli

Evita il backup di dati senza valore aziendale per ridurre i requisiti delle risorse di archiviazione per il tuo carico di lavoro.

Anti-pattern comuni:

- Non hai una strategia di backup per i tuoi dati.
- Esegui il backup di dati che possono essere facilmente ricreati.

Vantaggi dell'adozione di questa best practice: riduzione delle risorse di archiviazione necessarie per il carico di lavoro, oltre al relativo impatto ambientale, evitando il backup di dati non critici.

Livello di rischio associato se questa best practice non fosse adottata: medio

Guida all'implementazione

Evitando il backup di dati non necessari si possono ridurre i costi e le risorse di archiviazione utilizzate dal carico di lavoro. Esegui il backup solo dei dati che hanno un valore aziendale o sono considerati necessari per soddisfare i requisiti di conformità. Esamina le policy di backup ed escludi l'archiviazione temporanea che non offre valore in uno scenario di ripristino.

Passaggi dell'implementazione

- Implementa la policy di classificazione dei dati come illustrato in SUS04-BP01 Implementare una politica di classificazione dei dati.
- Sfrutta la criticità della classificazione dei dati e progetta una strategia di backup basata sull'obiettivo del tempo di ripristino (RTO) e sull'obiettivo del punto di ripristino (). RPO Evita il backup di dati non critici.
  - Escludi i dati che possono essere facilmente ricreati.
  - Escludi dati temporanei dai backup.
  - Escludi le copie locali dei dati, a meno che il tempo necessario per ripristinare i dati da una posizione comune non superi gli accordi sui livelli di servizio (). SLAs
- Usa una soluzione automatizzata o un servizio gestito per eseguire il backup di dati aziendali strategici.
  - AWS Backupè un servizio completamente gestito che semplifica la centralizzazione e l'automazione della protezione dei dati tra i AWS servizi, nel cloud e in locale. Per una guida pratica su come creare backup automatici utilizzando, AWS Backup consulta Well-Architected Labs - Testing Backup and Restore of Data.
  - Automatizza i backup e ottimizza i costi di backup per l'utilizzo di Amazon EFS. AWS Backup

Risorse

Best practice correlate:

- REL09-BP01 Identifica ed esegui il backup di tutti i dati di cui è necessario eseguire il backup o riproduci i dati dalle fonti
- REL09-BP03 Esegui automaticamente il backup dei dati
- REL13-BP02 Utilizza strategie di ripristino definite per soddisfare gli obiettivi di ripristino

Documenti correlati:

- Utilizzo AWS Backup per il backup e il ripristino dei EFS file system Amazon
- EBSIstantanee Amazon
- Utilizzo dei backup su Amazon Relational Database Service
- APNPartner: partner che possono aiutarti con il backup
- Marketplace AWS: prodotti che possono essere utilizzati per il backup
- Eseguire il backup di Amazon EFS
- Backup di Amazon FSx per Windows File Server
- Backup e ripristino per Amazon ElastiCache (RedisOSS)

Video correlati:

- AWS re:Invent 2023 - Strategie di backup e disaster recovery per una maggiore resilienza
- AWS re:Invent 2023 - Cosa c'è di nuovo con AWS Backup
- AWS re:Invent 2021 - Backup, disaster recovery e protezione dal ransomware con AWS

Esempi correlati:

- Well-Architected Lab: dati di backup

Hardware e servizi

Domanda

- SUS5 Come selezionate e utilizzate l'hardware e i servizi cloud nella vostra architettura per supportare i vostri obiettivi di sostenibilità?

SUS5 Come selezionate e utilizzate l'hardware e i servizi cloud nella vostra architettura per supportare i vostri obiettivi di sostenibilità?

Cerca opportunità per ridurre l'impatto dei carichi di lavoro in termini di sostenibilità apportando modifiche alle tue pratiche di gestione hardware. Riduci al minimo la quantità di hardware necessaria per il provisioning e l'implementazione e scegli l'hardware e i servizi più efficienti per il singolo carico di lavoro.

Best practice

- SUS05-BP01 Utilizza la quantità minima di hardware per soddisfare le tue esigenze
- SUS05-BP02 Usa i tipi di istanza con il minor impatto
- SUS05-BP03 Usa servizi gestiti
- SUS05-BP04 Ottimizza l'uso degli acceleratori di elaborazione basati su hardware

SUS05-BP01 Utilizza la quantità minima di hardware per soddisfare le tue esigenze

Usa la quantità minima di hardware per il tuo carico di lavoro per soddisfare in modo efficiente le tue esigenze aziendali.

Anti-pattern comuni:

- Non monitori l'utilizzo delle risorse.
- Nella tua architettura sono presenti risorse con un basso livello di utilizzo.
- Non analizzi l'uso di hardware statico per stabilire se deve essere ridimensionato.
- Non si stabiliscono obiettivi di utilizzo dell'hardware per l'infrastruttura di elaborazione in base al business. KPIs

Vantaggi dell'adozione di questa best practice: riduzione dell'impatto ambientale dei carichi di lavoro, risparmio di denaro e mantenimento dei benchmark delle prestazioni grazie al ridimensionamento corretto delle risorse cloud.

Livello di rischio associato se questa best practice non fosse adottata: medio

Guida all'implementazione

Seleziona con precisione la quantità di hardware richiesta dal tuo carico di lavoro per migliorare l'efficienza generale. Cloud AWS Offre la flessibilità necessaria per espandere o ridurre il numero di risorse in modo dinamico attraverso una varietà di meccanismi, ad esempio per soddisfare le variazioni della domanda. AWS Auto Scaling Fornisce inoltre APIse consente SDKs di modificare le risorse con il minimo sforzo. Usa queste funzionalità per apportare modifiche frequenti alle implementazioni dei carichi di lavoro. Inoltre, utilizza le linee guida sul corretto dimensionamento fornite dagli AWS strumenti per gestire in modo efficiente le risorse cloud e soddisfare le esigenze aziendali.

Passaggi dell'implementazione

- Scegli il tipo di istanza: scegli il giusto tipo di istanza così da soddisfare appieno le tue esigenze. Per scoprire come scegliere le istanze Amazon Elastic Compute Cloud e utilizzare meccanismi quali la selezione delle istanze basata sugli attributi, consulta le seguenti risorse:
  - Come faccio a scegliere il tipo di EC2 istanza Amazon appropriato per il mio carico di lavoro?
  - Selezione del tipo di istanza basata sugli attributi per Amazon Fleet. EC2
  - Crea un gruppo Auto Scaling utilizzando la selezione del tipo di istanza basata su attributi.
- Dimensiona usa piccoli incrementi per scalare carichi di lavoro variabili.
- Ricorri a più opzioni di acquisto di calcolo: bilancia flessibilità, scalabilità e risparmi sui costi delle istanze con più opzioni di acquisto di calcolo.
  - Le istanze Amazon EC2 On-Demand sono più adatte per carichi di lavoro nuovi, stateful e con picchi, che non possono essere flessibili in termini di tipo di istanza, ubicazione o orario.
  - Le istanze Amazon EC2 Spot sono un ottimo modo per integrare le altre opzioni per applicazioni flessibili e tolleranti ai guasti.
  - Sfrutta i Savings Plans per il calcolo per carichi di lavoro a stato costante che garantiscono la flessibilità in caso di cambiamento delle le tue esigenze (come zone di disponibilità, regioni, famiglie di istanze o tipi di istanze).
- Usa la diversità di istanze e zone di disponibilità: ottimizza la disponibilità delle applicazioni e sfrutta la capacità in eccesso diversificando istanze e zone di disponibilità.
- Istanze della dimensione giusta: utilizza i consigli sul dimensionamento corretto forniti dagli AWS strumenti per apportare modifiche al carico di lavoro. Per ulteriori informazioni, consulta Optimizing your cost with Rightsizing Recommendations e Right Sizing: Provisioning Instances to Match Workloads.
  - Utilizza i consigli sul corretto dimensionamento in o per identificare opportunità di dimensionamento corretto. AWS Cost Explorer AWS Compute Optimizer
- Negoziazione di accordi sui livelli di servizio (SLAs): negoziazione SLAs che consentano di ridurre temporaneamente la capacità mentre l'automazione impiega risorse sostitutive.

Risorse

Documenti correlati:

- Ottimizzazione dell' AWS infrastruttura per la sostenibilità, parte I: Elaborazione
- Selezione del tipo di istanza basata sugli attributi per Auto Scaling for Amazon Fleet EC2
- AWS Compute Optimizer Documentazione

- Operating Lambda: Performance optimization
- Documentazione su Auto Scaling

Video correlati:

- AWS re:Invent 2023 - Cosa c'è di nuovo con Amazon EC2
- AWS re:Invent 2023 - Risparmio intelligente: strategie di ottimizzazione dei costi di Amazon Elastic Compute Cloud
- AWS re:Invent 2022 - Ottimizzazione di Amazon Elastic Kubernetes Service per prestazioni e costi AWS
- AWS re:Invent 2023 - Elaborazione sostenibile: riduzione dei costi e delle emissioni di carbonio con AWS

SUS05-BP02 Usa i tipi di istanza con il minor impatto

Esegui un monitoraggio costante e usa nuovi tipi di istanza per sfruttare le migliorie in termini di efficienza energetica.

Anti-pattern comuni:

- Utilizzi una sola famiglia di istanze.
- Utilizzi solo istanze x86.
- È necessario specificare un tipo di istanza nella configurazione di Amazon EC2 Auto Scaling.
- AWS Le istanze vengono utilizzate in un modo per cui non sono state progettate (ad esempio, si utilizzano istanze ottimizzate per il calcolo per un carico di lavoro che richiede molta memoria).
- Non valuti regolarmente l'uso di nuovi tipi di istanza.
- Non si controllano i consigli forniti da strumenti di dimensionamento corretto, ad esempio. AWSAWS Compute Optimizer

Vantaggi dell'adozione di questa best practice: l'uso di istanze energeticamente efficienti e di dimensioni corrette ti consente di ridurre in modo considerevole l'impatto ambientale e i costi del carico di lavoro.

Livello di rischio associato se questa best practice non fosse adottata: medio

Guida all'implementazione

L'uso di istanze efficienti nel carico di lavoro cloud è fondamentale per ridurre l'utilizzo delle risorse e i costi. Monitora costantemente il rilascio di nuovi tipi di istanze e sfrutta le migliorie in tema di efficienza energetica, inclusi i tipi di istanze progettati per supportare carichi di lavoro specifici, come la formazione del machine learning, le inferenze e la transcodifica dei video.

Passaggi dell'implementazione

- Scopri e approfondisci i tipi di istanze: esplora e approfondisci i tipi di istanza in grado di ridurre l'impatto ambientale del carico di lavoro.
  - Abbonati a What's New with AWS per rimanere aggiornato up-to-date sulle ultime AWS tecnologie e istanze.
  - Scopri i diversi tipi di AWS istanze.
  - Scopri le istanze AWS basate su Graviton che offrono le migliori prestazioni per watt di consumo energetico su Amazon EC2 guardando RE:Invent 2020 - Approfondimento sulle istanze Amazon AWS basate sul processore Graviton2 e Approfondimento sulle istanze Graviton3 e Amazon EC2 C7g. AWS EC2
- Usa i tipi di istanza che comportano il minor impatto: pianifica la transizione del carico di lavoro a tipi di istanza caratterizzati dal minimo impatto.
  - Definisci un processo per valutare nuove funzionalità o istanze per il carico di lavoro. Sfrutta l'agilità del cloud per testare in modo semplice e rapido in che modo i nuovi tipi di istanza possono migliorare la sostenibilità ambientale del carico di lavoro. Utilizza metriche proxy per misurare la quantità di risorse necessarie per completare un'unità di lavoro.
  - Se possibile, modifica il carico di lavoro in modo che funzioni con diversi numeri e quantità di memoria per massimizzare la scelta del tipo di istanza. vCPUs
  - Valuta l'ipotesi di trasferire il carico di lavoro in istanze basate su Graviton per migliorare l'efficienza delle prestazioni del carico di lavoro. Per ulteriori informazioni sullo spostamento dei carichi di lavoro su AWS Graviton, consulta AWS Graviton Fast Start e Considerazioni sulla transizione dei carichi di lavoro a istanze Amazon Elastic Compute Cloud basate su Graviton AWS.
  - Prendi in considerazione la possibilità di selezionare l'opzione Graviton nell'utilizzo dei servizi gestiti. AWSAWS
  - Esegui la migrazione del carico di lavoro nelle regioni che offrono istanze con il minor impatto in termini di sostenibilità e che contemporaneamente soddisfano i requisiti aziendali.

  - Per i carichi di lavoro di machine learning, sfrutta hardware appositamente progettato e specifico per il tuo carico di lavoro come AWS Trainium, Inferentia e Amazon.AWS EC2 DL1 AWS Le istanze Inferentia come le istanze Inf2 offrono prestazioni per watt migliori fino al 50% rispetto alle istanze Amazon comparabili. EC2
  - Usa Amazon SageMaker Inference Recommender per un endpoint di inferenza ML di dimensioni corrette.
  - Per carichi di lavoro con picchi (carichi di lavoro con requisiti non frequenti di capacità aggiuntiva), utilizza istanze a prestazioni espandibili.
  - Per carichi di lavoro stateless e con tolleranza ai guasti, usa Amazon EC2 Spot Instances per aumentare l'utilizzo complessivo del cloud e ridurre l'impatto sulla sostenibilità delle risorse inutilizzate.
- Esegui e ottimizza: esegui e ottimizza l'istanza del carico di lavoro.
  - Per carichi di lavoro temporanei, valuta i CloudWatch parametri di Amazon dell'istanza, ad esempio `CPUUtilization` per identificare se l'istanza è inattiva o sottoutilizzata.
  - Per carichi di lavoro stabili, controlla gli strumenti di dimensionamento AWS corretto, ad esempio a intervalli regolari, per identificare le opportunità di ottimizzazione e dimensionamento corretto delle istanze AWS Compute Optimizer. Per ulteriori esempi e consigli, consulta i seguenti lab:
    - Well-Architected Lab: raccomandazioni per il ridimensionamento corretto
    - Well-Architected Lab: ridimensionamento corretto con Compute Optimizer
    - Well-Architected Lab - Ottimizza i modelli hardware e rispetta la sostenibilità KPIs

Risorse

Documenti correlati:

- Ottimizzazione dell' AWS infrastruttura per la sostenibilità, parte I: Elaborazione
- AWS Gravitone
- Amazon EC2 DL1
- Flotte EC2 di prenotazione della capacità di Amazon
- Flotta Amazon EC2 Spot
- Funzioni: configurazione della funzione Lambda
- Selezione del tipo di istanza basata sugli attributi per Amazon Fleet EC2
- Building Sustainable, Efficient, and Cost-Optimized Applications on AWS

- How the Contino Sustainability Dashboard Helps Customers Optimize Their Carbon Footprint

Video correlati:

- AWS re:Invent 2023 - AWS Graviton: il miglior rapporto prezzo/prestazioni per i tuoi carichi di lavoro AWS
- AWS re:Invent 2023 - Nuove funzionalità di intelligenza artificiale generativa di Amazon Elastic Compute Cloud in AWS Management Console
- AWS re:Invent 2023 = Novità di Amazon Elastic Compute Cloud
- AWS re:Invent 2023 - Risparmio intelligente: strategie di ottimizzazione dei costi di Amazon Elastic Compute Cloud
- AWS re:Invent 2021 - Approfondimento sulle istanze AWS Graviton3 e Amazon C7g EC2
- AWS re:Invent 2022 - Crea un ambiente di elaborazione efficiente in termini di costi, energia e risorse

Esempi correlati:

- Soluzione: linee guida per l'ottimizzazione dei carichi di lavoro di deep learning per la sostenibilità su AWS
- Migrating Amazon Relational Database Service Databases to Graviton

SUS05-BP03 Usa servizi gestiti

Usa i servizi gestiti per operare in modo più efficiente nel cloud.

Anti-pattern comuni:

- Utilizzi EC2 istanze Amazon a basso utilizzo per eseguire le tue applicazioni.
- Il tuo team interno gestisce solo il carico di lavoro, senza tempo per focalizzarsi sull'innovazione o sulle semplificazioni.
- Implementi e mantieni tecnologie per attività che possono essere eseguite in modo più efficiente sui servizi gestiti.

Vantaggi dell'adozione di questa best practice:

- L'uso dei servizi gestiti sposta la responsabilità verso AWS, che dispone di informazioni su milioni di clienti che possono contribuire a promuovere nuove innovazioni ed efficienze.
- Il servizio gestito distribuisce l'impatto ambientale del servizio su molti utenti a causa dei piani di controllo (control-plane) multi-tenet.

Livello di rischio associato se questa best practice non fosse adottata: medio

Guida all'implementazione

I servizi gestiti trasferiscono la AWS responsabilità al mantenimento di un elevato utilizzo e all'ottimizzazione della sostenibilità dell'hardware distribuito. I servizi gestiti eliminano anche l'onere operativo e amministrativo legato alla manutenzione di un servizio, consentendo al tuo team di avere più tempo e di concentrarsi sull'innovazione.

Esamina il carico di lavoro per identificare i componenti che possono essere sostituiti dai AWS servizi gestiti. Ad esempio, Amazon RDS, Amazon Redshift e Amazon ElastiCache forniscono un servizio di database gestito. Amazon AthenaEMR, Amazon e Amazon OpenSearch Service forniscono un servizio di analisi gestito.

Passaggi dell'implementazione

1. Esegui l'inventario del carico di lavoro: esegui un inventario del tuo carico di lavoro in relazione a servizi e componenti.
2. Identifica i candidati: procedi a valutare e identificare i componenti sostituibili dai servizi gestiti. Ecco alcuni esempi in cui potresti prendere in considerazione l'uso di un servizio gestito:

Attività	Cosa usare su AWS
Ospitare un database	Utilizza istanze gestite di Amazon Relational Database Service (RDSAmazon) invece di mantenere le tue istanze Amazon su RDS Amazon Elastic Compute Cloud EC2 (Amazon).
Ospitare il carico di lavoro di un container	Utilizza AWS Fargate, invece di implementare un'infrastruttura di container proprietaria.

Attività	Cosa usare su AWS
Ospitare applicazioni Web	Usa AWS Amplify Hosting come servizio CI/CD e di hosting completamente gestito per siti Web statici e app Web con rendering lato server.

3. Crea un piano di migrazione: individua le dipendenze e crea un piano di migrazione. Aggiorna runbook e playbook di conseguenza
   - AWS Application Discovery Service raccoglie e presenta automaticamente informazioni dettagliate sulle dipendenze e sull'utilizzo delle applicazioni per aiutarti a prendere decisioni più informate durante la pianificazione della migrazione.
4. Esegui i test: testa il servizio prima di migrare al servizio gestito.
5. Sostituisci i servizi in hosting autonomo: utilizza il tuo piano di migrazione per sostituire i servizi in hosting autonomo con servizi gestiti.
6. Monitora e modifica: monitora costantemente il servizio al termine della migrazione per apportare le modifiche richieste e ottimizzare il servizio.

Risorse

Documenti correlati:

- Cloud AWS Prodotti
- AWS Calcolatore del costo totale di proprietà (TCO)
- Amazon DocumentDB
- Servizio Amazon Elastic Kubernetes () EKS
- Streaming gestito da Amazon per Apache Kafka (Amazon) MSK

Video correlati:

- AWS re:Invent 2021 - Operazioni cloud su larga scala con AWS Managed Services
- AWS re:Invent 2023 - Le migliori pratiche per operare su AWS

SUS05-BP04 Ottimizza l'uso degli acceleratori di elaborazione basati su hardware

Ottimizza l'uso delle istanze a calcolo accelerato per ridurre i requisiti dell'infrastruttura fisica del carico di lavoro.

Anti-pattern comuni:

- Non stai monitorando GPU l'utilizzo.
- Utilizzo di un'istanza per uso generico per il carico di lavoro quando un'istanza appositamente sviluppata potrebbe offrire prestazioni più elevate, costi inferiori e migliori prestazioni per watt.
- Stai utilizzando acceleratori di calcolo basati su hardware per attività in cui sono più efficienti utilizzando alternative basate su hardware. CPU

Vantaggi dell'adozione di questa best practice: ottimizzando l'uso degli acceleratori basati su hardware, è possibile ridurre le richieste di infrastruttura fisica del carico di lavoro.

Livello di rischio associato se questa best practice non fosse adottata: medio

Guida all'implementazione

Se hai bisogno di capacità di elaborazione elevate, puoi trarre vantaggio dall'utilizzo di istanze di elaborazione accelerata, che forniscono l'accesso ad acceleratori di elaborazione basati su hardware come le unità di elaborazione grafica () e gli array di porte programmabili sul campo (). GPUs FPGAs Questi acceleratori hardware eseguono determinate funzioni, come l'elaborazione grafica o la corrispondenza dei modelli di dati, in modo più efficiente rispetto alle alternative basate. CPU Molti carichi di lavoro accelerati, come il rendering grafico, la transcodifica e il machine learning, sono altamente variabili in termini di utilizzo di risorse. Mantieni in esecuzione questo tipo di hardware solo per il tempo necessario e disattivalo automaticamente quando non serve per ridurre la quantità di risorse utilizzate.

Passaggi dell'implementazione

- Identifica le istanza a calcolo accelerato in grado di soddisfare i tuoi requisiti.
- Per i carichi di lavoro di machine learning, sfrutta l'hardware appositamente progettato e specifico per il tuo carico di lavoro, come AWS Trainium, Inferentia e Amazon.AWS EC2 DL1 AWS Le istanze Inferentia come le istanze Inf2 offrono prestazioni per watt migliori fino al 50% rispetto alle istanze Amazon comparabili. EC2

- Raccogli i parametri di utilizzo delle istanze a calcolo accelerato. Ad esempio, puoi utilizzare l' CloudWatch agente per raccogliere metriche come `utilization_gpu` e `utilization_memory` per le tue, GPUs come mostrato in Collect NVIDIA GPU metrics with Amazon. CloudWatch
- Ottimizza il codice, il funzionamento della rete e le impostazioni degli acceleratori hardware per garantire il pieno utilizzo dell'hardware sottostante.
  - Ottimizza le impostazioni GPU
  - GPUMonitoraggio e ottimizzazione nel deep learning AMI
  - Ottimizzazione dell'I/O per l'ottimizzazione GPU delle prestazioni della formazione di deep learning in Amazon SageMaker
- Utilizza le librerie e i driver più recenti ad alte prestazioni. GPU
- Usa l'automazione per rilasciare GPU istanze quando non sono in uso.

Risorse

Documenti correlati:

- Calcolo accelerato
- Let's Architect! Architecting with custom chips and accelerators
- Come faccio a scegliere il tipo di EC2 istanza Amazon appropriato per il mio carico di lavoro?
- EC2VT1Istanze Amazon
- Scegli l'acceleratore di intelligenza artificiale e la compilazione di modelli migliori per l'inferenza della visione artificiale con Amazon SageMaker

Video correlati:

- AWS re:Invent 2021 - Come selezionare le EC2 GPU istanze Amazon per il deep learning
- AWS Colloqui tecnici online - Implementazione di inferenze di deep learning a costi contenuti
- AWS re:Invent 2023 - AI all'avanguardia con e AWS NVIDIA
- AWS re:Invent 2022 - [!] NEW LAUNCH Presentazione delle AWS istanze Amazon Inf2 basate su Inferentia2 EC2
- AWS re:Invent 2022 - Accelera il deep learning e innova più velocemente con AWS Trainium
- AWS re:Invent 2022 - Apprendimento approfondito con: dalla formazione all'implementazione AWS NVIDIA

Processo e cultura

Domanda

- SUS6 In che modo i vostri processi organizzativi supportano i vostri obiettivi di sostenibilità?

SUS6 In che modo i vostri processi organizzativi supportano i vostri obiettivi di sostenibilità?

Cerca opportunità per ridurre l'impatto di sostenibilità apportando modifiche alle tue prassi di sviluppo, test e implementazione.

Best practice

- SUS06-BP01 Adottare metodi in grado di introdurre rapidamente miglioramenti della sostenibilità
- SUS06-BP02 Mantieni il tuo carico di lavoro up-to-date
- SUS06-BP03 Aumentare l'utilizzo degli ambienti di costruzione
- SUS06-BP04 Usa farm di dispositivi gestiti per i test

SUS06-BP01 Adottare metodi in grado di introdurre rapidamente miglioramenti della sostenibilità

Adotta metodi e processi per convalidare migliorie potenziali, ridurre i costi legati ai test e offrire piccole migliorie.

Anti-pattern comuni:

- Analizzare l'applicazione rispetto alla sostenibilità è un'attività che viene eseguita solo una volta, all'inizio di un progetto.
- Il tuo carico di lavoro non è aggiornato, poiché il processo di rilascio è troppo complesso per introdurre modifiche minori per l'efficienza delle risorse.
- Non hai meccanismi per migliorare il tuo carico di lavoro in termini di sostenibilità.

Vantaggi dell'adozione di questa best practice: la definizione di un processo per l'introduzione e il monitoraggio dei miglioramenti della sostenibilità consente di adottare in modo continuo nuove funzionalità e funzioni, risolvere i problemi e migliorare l'efficienza del carico di lavoro.

Livello di rischio associato se questa best practice non fosse adottata: medio

Guida all'implementazione

Testa e convalida potenziali miglioramenti all'impatto sulla sostenibilità prima di implementarli in produzione. Tieni in considerazione il costo dei test quando calcoli il potenziale vantaggio futuro di un miglioramento. Sviluppa metodi di test a basso costo per consentire la distribuzione di piccoli miglioramenti.

Passaggi dell'implementazione

- Analizza e comunica i tuoi obiettivi di sostenibilità organizzativa: esamina i tuoi obiettivi di sostenibilità organizzativa, come la riduzione delle emissioni di carbonio o la gestione delle risorse idriche. Traduci questi obiettivi in requisiti di sostenibilità per i carichi di lavoro del cloud. Comunica questi requisiti alle principali parti interessate.
- Aggiungi i requisiti di sostenibilità al tuo backlog: aggiungi i requisiti relativi al miglioramento della sostenibilità al tuo backlog di sviluppo.
- Itera e migliora: utilizza un processo di miglioramento iterativo per identificare, valutare, assegnare priorità, testare e implementare questi miglioramenti.
- Test utilizzando un prodotto minimo valido (MVP): Sviluppa e testa potenziali miglioramenti utilizzando i componenti rappresentativi minimi possibili per ridurre i costi e l'impatto ambientale dei test.
- Semplifica il processo: migliora e semplifica continuamente i tuoi processi di sviluppo. Ad esempio, automatizza il processo di distribuzione del software con pipeline di distribuzione e integrazione continue (CI/CD) per testare e implementare migliorie potenziali per ridurre il livello di impegno e gli errori causati da processi manuali.
- Gestisci formazione e sensibilizzazione: organizza programmi di formazione per i membri del tuo team per sensibilizzarli in merito alla sostenibilità e sull'impatto delle loro attività sugli obiettivi di sostenibilità dell'organizzazione.
- Valuta e modifica: valuta in modo costante l'impatto delle migliorie e apporta gli adeguamenti richiesti.

Risorse

Documenti correlati:

- AWS abilita soluzioni di sostenibilità
- Pratiche di sviluppo agili e scalabili basate su AWS CodeCommit

Video correlati:

- AWS re:Invent 2023 - Architettura sostenibile: passato, presente e futuro
- AWS re:Invent 2022 - Fornire architetture sostenibili e ad alte prestazioni
- AWS re:Invent 2022 - Progettare in modo sostenibile e ridurre l'impronta di carbonio AWS
- AWS re:Invent 2022 - Sostenibilità nelle infrastrutture globali AWS
- AWS re:Invent 2023 - Cosa c'è di nuovo con l'osservabilità e le operazioni AWS

Esempi correlati:

- Well-Architected Lab: trasformare i report su costi e utilizzo in report sull'efficienza

SUS06-BP02 Mantieni il tuo carico di lavoro up-to-date

Mantieni il carico up-to-date di lavoro per adottare funzionalità efficienti, rimuovere problemi e migliorare l'efficienza complessiva del carico di lavoro.

Anti-pattern comuni:

- Si ritiene che l'architettura corrente diventi statica e non venga aggiornata nel corso del tempo.
- Non si dispone di sistemi né si esegue regolarmente una valutazione per la compatibilità di software e pacchetti aggiornati con il carico di lavoro.

Vantaggi dell'adozione di questa best practice: la definizione di un processo per garantire il costante aggiornamento del carico di lavoro ti consentirà di adottare nuove caratteristiche e funzionalità, risolvere i problemi e migliorare l'efficienza del carico di lavoro.

Livello di rischio associato se questa best practice non fosse adottata: basso

Guida all'implementazione

Sistemi operativi, runtime, middleware (software intermediario), librerie e applicazioni aggiornati possono incidere sull'efficienza dei carichi di lavoro e facilitano l'adozione delle tecnologie più efficienti. Il software aggiornato potrebbe anche includere funzionalità per misurare in modo più accurato l'impatto in termini di sostenibilità del carico di lavoro, poiché i fornitori offrono caratteristiche per raggiungere i propri obiettivi di sostenibilità. Adotta una cadenza regolare per aggiornare il tuo carico di lavoro con le ultime funzionalità e i rilasci più recenti.

Passaggi dell'implementazione

- Definisci un processo: serviti di un processo e una pianificazione per valutare nuove funzionalità o istanze per il carico di lavoro. Sfrutta l'agilità del cloud per testare in modo semplice e rapido il modo in cui le nuove funzionalità possono migliorare il carico di lavoro nei seguenti ambiti:
  - Riduzione dell'impatto a livello di sostenibilità.
  - Raggiungimento di maggiore efficienza in termini di prestazioni.
  - Eliminazione delle barriere finalizzata a un miglioramento pianificato.
  - Miglioramento della capacità di misurare e gestire l'impatto a livello di sostenibilità.
- Esegui l'inventario: redigi l'inventario del software e dell'architettura del carico di lavoro e identifica i componenti che richiedono un aggiornamento.
  - Puoi utilizzare AWS Systems Manager Inventory per raccogliere i metadati del sistema operativo (OS), delle applicazioni e delle istanze dalle tue EC2 istanze Amazon e capire rapidamente quali istanze eseguono il software e le configurazioni richieste dalla tua politica software e quali istanze devono essere aggiornate.
- Apprendi la procedura di aggiornamento: scopri come aggiornare i componenti del carico di lavoro.

Componente del carico di lavoro	Come aggiornare
Immagini della macchina	Usa EC2Image Builder per gestire gli aggiornamenti alle immagini dei server Amazon Machine Images (AMIs) per Linux o Windows.
Immagini di container	Usa Amazon Elastic Container Registry (AmazonECR) con la tua pipeline esistente per gestire le immagini di Amazon Elastic Container Service (AmazonECS).
AWS Lambda	AWS Lambda include funzionalità di gestione delle versioni.

- Utilizza l'automazione: usa l'automazione degli aggiornamenti per ridurre il livello di impegno per implementare le nuove funzionalità e limitare gli errori causati dai processi manuali.
  - Puoi usare CI/CD per aggiornare AMIs automaticamente le immagini dei container e altri elementi relativi alla tua applicazione cloud.

- È possibile utilizzare strumenti come Patch Manager di AWS Systems Manager per automatizzare il processo di aggiornamento del sistema e pianificare l'attività utilizzando le finestre di manutenzione di AWS Systems Manager.

Risorse

Documenti correlati:

- AWS Centro di architettura
- Cosa c'è di nuovo con AWS
- AWS Strumenti per developer

Video correlati:

- AWS re:Invent 2022 - Ottimizza i tuoi AWS carichi di lavoro con una guida basata sulle migliori pratiche
- All Things Patch: AWS Systems Manager

Esempi correlati:

- Well-Architected Labs: inventario e gestione delle patch
- Laboratorio: AWS Systems Manager

SUS06-BP03 Aumentare l'utilizzo degli ambienti di costruzione

Aumenta l'uso delle risorse per sviluppare, testare e creare i tuoi carichi di lavoro.

Anti-pattern comuni:

- Esegui il provisioning manuale o interrompi i tuoi ambienti di sviluppo.
- Fai in modo che i tuoi ambienti di sviluppo siano in esecuzione indipendentemente dalle attività di test, creazione o rilascio (ad esempio, eseguire un ambiente al di fuori dell'orario di lavoro dei membri del tuo team di sviluppo).
- Esegui un provisioning eccessivo delle tue risorse per gli ambienti di creazione.

Vantaggi dell'adozione di questa best practice: l'aumento dell'utilizzo degli ambienti di compilazione migliora l'efficienza complessiva del carico di lavoro in cloud, allocando al contempo le risorse agli sviluppatori per sviluppo, test e compilazione ottimali.

Livello di rischio associato se questa best practice non fosse adottata: basso

Guida all'implementazione

Usa l'automazione e attiva infrastructure-as-code gli ambienti di costruzione quando necessario e smontali quando non vengono utilizzati. Un modello comune consiste nel pianificare periodi di disponibilità che coincidano con l'orario di lavoro dei membri del team incaricati dello sviluppo. Gli ambienti di test devono essere molto simili alla configurazione di produzione. Tuttavia, cerca opportunità per utilizzare tipi di istanze con capacità burst, istanze Amazon EC2 Spot, servizi di database a scalabilità automatica, contenitori e tecnologie serverless per allineare la capacità di sviluppo e test all'uso. Limita i volumi di dati per soddisfare solo i requisiti di test. Se usi i dati di produzione per i test, rifletti sulla possibilità di condividere i dati di produzione invece di spostarli.

Passaggi dell'implementazione

- Utilizza il modello Infrastructure as code: usa il modello Infrastructure as code per eseguire il provisioning dei tuoi ambienti di sviluppo.
- Utilizza l'automazione: usa l'automazione per gestire il ciclo di vita degli ambienti di sviluppo e test e massimizzare l'efficienza delle tue risorse di sviluppo.
- Massimizza l'utilizzo: utilizza strategie per ottimizzare l'utilizzo degli ambienti di sviluppo e test.
  - Utilizza ambienti rappresentativi minimi realizzabili per lo sviluppo e il test di potenziali miglioramenti.
  - Utilizza tecnologie serverless, se possibile.
  - Utilizza istanze on-demand per integrare i dispositivi per gli sviluppatori.
  - Utilizza i tipi di istanze con capacità di espansione, istanze spot e altre tecnologie per allineare la capacità di compilazione all'uso.
  - Adotta servizi cloud nativi per un accesso sicuro agli shell (interprete di comandi) delle istanze invece di implementare parchi istanze di host bastioni.
  - Dimensiona automaticamente le tue risorse di sviluppo in base alle tue attività.

Risorse

Documenti correlati:

- AWS Systems Manager Gestore delle sessioni
- Istanze a prestazioni Amazon EC2 Burstable
- Che cos'è AWS CloudFormation?
- What is AWS CodeBuild?
- Instance Scheduler attivo AWS

Video correlati:

- AWS re:Invent 2023 - Integrazione e distribuzione continue per AWS

SUS06-BP04 Usa farm di dispositivi gestiti per i test

Usa device farm gestite per testare in maniera efficiente una nuova funzionalità su un set rappresentativo di hardware.

Anti-pattern comuni:

- Testa e implementi manualmente la tua applicazione su singoli dispositivi fisici.
- Non utilizzi il servizio di test delle app per testare e interagire con le tue app (ad esempio, Android, iOS e app Web) su dispositivi fisici reali.

Vantaggi dell'adozione di questa best practice: l'utilizzo di farm di dispositivi gestiti per il test delle applicazioni abilitate al cloud offre una serie di vantaggi.

- Offrono funzionalità più efficienti per testare le applicazioni su un'ampia gamma di dispositivi.
- Eliminano la necessità di un'infrastruttura in-house per i test.
- Offrono diverse tipologie di dispositivi, tra cui hardware di generazioni precedenti e meno diffuso, eliminando così la necessità di aggiornamenti non necessari dei dispositivi.

Livello di rischio associato se questa best practice non fosse adottata: basso

Guida all'implementazione

L'uso di device farm gestite può aiutarti a semplificare il processo di test per le nuove funzionalità su un gruppo rappresentativo di hardware. Le device farm gestite offrono diversi tipi di dispositivi, inclusi hardware meno diffusi e di generazioni precedenti, ed evitano l'impatto sulla sostenibilità dei clienti dovuti ad aggiornamenti dei dispositivi non necessari.

Passaggi dell'implementazione

- Definisci i requisiti di test: definisci i requisiti di test ed esegui la pianificazione (come tipo di test, sistemi operativi e programma di test).
  - Puoi usare Amazon CloudWatch RUM per raccogliere e analizzare dati lato client e definire il tuo piano di test.
- Seleziona una device farm gestita: scegli una device farm gestita in grado di supportare i tuoi requisiti di test. Ad esempio, puoi utilizzare AWS Device Farm per testare e analizzare l'impatto delle modifiche su un set di hardware rappresentativo.
- Utilizza l'automazione: usa automazione e integrazione continua/l'implementazione continua (CI/CD) per pianificare ed eseguire i test.
  - Integrazione di AWS Device Farm con la tua pipeline CI/CD per eseguire test Selenium su più browser
  - Creazione e test di app iOS AWS DevOps e iPad OS con servizi mobili
- Rivedi e modifica: esamina sempre i risultati dei test e apporta le migliorie richieste.

Risorse

Documenti correlati:

- AWS Device Farm elenco dei dispositivi
- Visualizzazione del CloudWatch RUM pannello di controllo

Video correlati:

- AWS re:Invent 2023 - Migliora la qualità delle tue app per dispositivi mobili e web con Device Farm AWS
- AWS re:Invent 2021 - Ottimizza le applicazioni attraverso approfondimenti sugli utenti finali con Amazon CloudWatch RUM

Esempi correlati:

- AWS Device Farm App di esempio per Android
- AWS Device Farm App di esempio per iOS
- Test Appium Web per AWS Device Farm

Note

I clienti sono responsabili della propria valutazione indipendente delle informazioni contenute nel presente documento. Questo documento: (a) è solo a scopo informativo, (b) rappresenta le offerte e le pratiche attuali di AWS prodotti, che sono soggette a modifiche senza preavviso, e (c) non crea alcun impegno o assicurazione da parte dei suoi affiliati, AWS fornitori o licenzianti. AWS i prodotti o i servizi sono forniti «così come sono» senza garanzie, dichiarazioni o condizioni di alcun tipo, esplicite o implicite. Le responsabilità e le responsabilità dei AWS propri clienti sono regolate da AWS accordi e il presente documento non fa parte di, né modifica, alcun accordo tra AWS e i suoi clienti.

AWS Glossario

Per la AWS terminologia più recente, consultate il AWS glossario nella sezione Reference.Glossario AWS

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.