Canone di architettura AWS

Luglio 2020

This paper has been archived.

The latest version is now available at:

https://docs.aws.amazon.com/it_it/wellarchitected/latest/framework/welcome.html

Il presente documento descrive il Canone di architettura AWS, che ti permette di revisionare e migliorare le tue architetture basate sul cloud e comprendere meglio l'impatto aziendale delle decisioni di progettazione. In questo documento affrontiamo i principi generali di progettazione, best practice e linee guida specifiche in cinque aree concettuali che definiamo i *pilastri* del canone di architettura.

Avvisi

I clienti sono responsabili della propria valutazione autonoma delle informazioni contenute in questo documento. Questo documento: (a) è solo a scopo informativo, (b) mostra le offerte e le pratiche attuali dei prodotti AWS soggette a modifiche senza preavviso, e (c) non crea alcun impegno o garanzia da parte di AWS e dei suoi affiliati, fornitori o licenziatari. I prodotti o servizi AWS sono forniti "così come sono" senza garanzie, dichiarazioni o condizioni di alcun tipo, sia esplicite che implicite. Le responsabilità di AWS nei confronti dei propri clienti sono definite dai contratti AWS e il presente documento non costituisce parte né modifica qualsivoglia contratto tra AWS e i suoi clienti.

Introduzione

Il Canone di architettura AWS aiuta a comprendere i pro e i contro delle decisioni prese durante la progettazione di sistemi in AWS. Utilizzando il Canone, scoprirai le best practice architetturali per progettare e gestire sistemi affidabili, sicuri, efficienti e convenienti nel cloud. Permette di misurare in modo coerente le architetture rispetto alle best practice e identificare le aree da migliorare. Il processo di revisione di un'architettura consiste in una conversazione costruttiva sulle decisioni relative all'architettura e non è un meccanismo di audit. Disporre di sistemi ben architettati aumenta notevolmente la probabilità di successo aziendale.

AWS Solutions Architects ha anni di esperienza nell'architettura di soluzioni in un'ampia gamma di business e casi di utilizzo. 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 Canone di architettura AWS documenta un insieme di domande fondamentali per capire se un'architettura specifica si allinea bene con le best practice del cloud. Il canone 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 AWS continua a evolversi e noi continuiamo a imparare di più dal lavoro che svolgiamo con i nostri clienti, continueremo a ridefinire la definizione di canone di architettura.

Questo canone è rivolto a chi svolge ruoli tecnologici, ad esempio ai Chief Technology Officer (CTO), ai progettisti, agli sviluppatori e ai membri dei team operativi. Descrive le best practice e le strategie AWS da usare per la progettazione e il funzionamento di un carico di lavoro cloud, e fornisce collegamenti a ulteriori dettagli di implementazione e pattern architetturali. Per maggiori informazioni consulta la home page del Canone di Architettura AWS.

AWS offre anche un servizio gratuito di revisione dei carichi di lavoro. AWS Well-Architected Tool (AWS WA Tool) è un servizio cloud che fornisce un approccio coerente per la revisione e la valutazione della tua architettura tramite il Canone di architettura AWS. AWS WA Tool fornisce raccomandazioni per rendere i tuoi carichi di lavoro più affidabili, sicuri, efficienti e convenienti.

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 partner APN AWS Partner Network selezionati, che sono membri del programma AWS Well-Architected Partner. I suddetti partner APN hanno una conoscenza approfondita di AWS e possono aiutarti nella revisione e nel miglioramento dei tuoi carichi di lavoro.

Definizioni

Tutti i giorni esperti AWS supportano i clienti nella progettazione di sistemi di architettura per sfruttare le best practice nel cloud. Ti aiutiamo a trovare i compromessi relativi all'architettura nel processo di evoluzione dei tuoi progetti. Quando distribuisci questi sistemi in ambienti live, analizziamo le prestazioni di questi sistemi e le conseguenze dei suddetti compromessi.

Sulla base di quello che abbiamo imparato, abbiamo creato il Canone di architettura AWS, che fornisce un insieme coerente di best practice per i clienti e i partner per valutare le architetture, e fornisce un insieme di domande che puoi utilizzare per valutare se l'architettura è ben allineata alle best practice AWS.

Il Canone di architettura AWS si basa su cinque pilastri: eccellenza operativa, sicurezza, affidabilità, efficienza delle prestazioni e ottimizzazione dei costi.

Tabella 1. I pilastri del Canone di architettura AWS

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 continuamente i processi e le procedure di supporto per offrire valore aggiunto.
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.
Affidabilità	La capacità di un carico di lavoro di svolgere la funzione prevista in modo corretto e coerente quando previsto, ciò include la possibilità di utilizzare e testare il carico di lavoro durante il suo ciclo di vita totale.
Efficienza delle prestazioni	L'abilità di utilizzare in modo efficiente le risorse di elaborazione per soddisfare i requisiti del sistema e conservare tale efficienza a seconda dei cambiamenti della domanda e dell'evoluzione delle tecnologie.
Ottimizzazione dei costi	Capacità di eseguire sistemi per fornire valore aziendale al minor prezzo possibile

Nel Canone di architettura AWS usiamo questi termini

- Un **componente** è un codice, una configurazione e le risorse AWS 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. Il carico di lavoro, normalmente, è il livello di dettaglio comunicato dai leader aziendali e della tecnologia.
- **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).
- Pensiamo a un'**architettura** come al modo in cui in componenti operano insieme in un carico di lavoro. Il modo di comunicare e di interagire dei componenti è spesso l'aspetto principale dei diagrammi architetturali.
- Nell'ambito di un'organizzazione **il portfolio delle tecnologie** rappresenta l'insieme di carichi di lavoro necessari affinché l'azienda possa essere operativa.

Quando progetti l'architettura dei carichi di lavoro, devi trovare dei compromessi tra i pilastri su cui si regge il tuo contesto aziendale. Questo tipo di decisioni aziendali deve essere alla base delle tue priorità ingegneristiche. 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. Nelle soluzioni di e-commerce, le prestazioni possono avere un impatto sui profitti e sulla propensione all'acquisto da parte dei clienti. L'eccellenza in ambito di sicurezza e operatività generalmente non viene sacrificata rispetto agli altri pilastri.

Architettura

Negli ambienti in locale, 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 i team usano TOGAF o il Framework di Zachman come parte delle competenze architetturali aziendali.

Noi di AWS, preferiamo distribuire le competenze tra i team, invece di centralizzarle in un unico team. 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. Innanzitutto, abbiamo le *pratiche* [1] che hanno lo scopo di permettere a ogni team di possedere tali competenze e ci serviamo di esperti che garantiscano che i team adottino standard più severi di quelli che devono rispettare. In secondo luogo, implementiamo *meccanismi* [2] che eseguono controlli automatizzati per garantire che gli standard vengano rispettati. L'approccio distribui-

[1] Modalità per eseguire attività, processi, standard e norme accettate.
[2] «Le buone intenzioni non bastano mai, per avere successo servono buoni meccanismi» Jeff Bezos. Questo significa sostituire gli sforzi umani con meccanismi (spesso automatizzati) che verificano la conformità alle regole e ai processi.

to è supportato dai principi di leadership di Amazon, e stabilisce una cultura tra tutti i ruoli che *lavorano a ritroso* [3] a partire dalle necessità del cliente. 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 competenze o i team esistenti ad alzare il livello, abilitiamo l'accesso a una community virtuale di ingegneri responsabili che possono eseguire la revisione dei loro progetti e aiutarli a comprendere le best practice di AWS. La community di ingegneri responsabili 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.

Le best practice AWS sono il risultato della nostra esperienza nell'esecuzione di migliaia di sistemi su Internet. Preferiamo utilizzare i dati per definire le best practice, ma ci serviamo anche di esperti in materia come ingegneri responsabili per stabilire le best practice. Quando gli ingegneri responsabili vedono emergere nuove best practice lavorano con la community per garantire 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. 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 Architecture e i team di ingegneria interni. Well-Architected è un meccanismo scalabile che ti permette di sfruttare queste nozioni.

Seguendo l'approccio della community di ingegneri responsabili 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 della tecnologia (come i CTO o i manager dello sviluppo) che eseguono revisioni Well-Architected tra tutti i carichi di lavoro ti permettono di comprendere più a fondo i rischi relativi al portfolio delle tecnologie. Tramite questo approccio puoi identificare dei temi tra i team che la tua organizzazione può affrontare tramite meccanismi, formazione o lunchtime talks in cui gli ingegneri responsabili possono condividere le loro idee su aree specifiche con diversi team.

Principi generali di progettazione

Il Canone di architettura identifica una serie di principi generali per facilitare la corretta progettazione nel cloud:

- **Smetti di ipotizzare quali siano le tue esigenze di capacità**: Non formulare più ipotesi sui requisiti di capacità della tua infrastruttura. Quando prendi decisioni re-

[3] 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.

lative alla capacità prima della distribuzione di un sistema, potresti 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 ridimensionare 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 in locale.
- **Automatizza per facilitare la sperimentazione dell'architettura**: L'automazione ti permette di creare e replicare i tuoi sistemi a basso costo e di evitare le spese del lavoro manuale. Puoi tenere traccia delle modifiche all'automazione, effettuare l'audit dell'impatto e tornare ai parametri precedenti, se necessario.
- **Consenti le architetture evoluzionistiche**: Consenti 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. Quando un'azienda e il suo contesto continuano a cambiare, le decisioni iniziali potrebbero ostacolare la capacità del sistema di soddisfare i requisiti aziendali in cambiamento. 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 tuo 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**: Esegui test sulle prestazioni della tua architettura e dei tuoi 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 cinque pilastri del canone

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 cinque pilastri di eccellenza operativa, sicurezza, affidabilità, efficienza delle prestazioni e ottimizzazione dei costi, 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.

Eccellenza operativa

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

Principi di progettazione

Esistono cinque principi di progettazione per eccellenza operativa nel cloud:

- **Esegui le operazioni come codice**: Nel cloud ti è possibile applicare la medesima disciplina di progettazione che utilizzi per il codice dell'applicazione a tutto il tuo ambiente. Puoi definire l'intero carico di lavoro (applicazioni, infrastruttura) come codice e aggiornarlo con il codice. Puoi implementare le tue procedure operative come codice e automatizzarne l'esecuzione attivandole in risposta agli eventi. Eseguendo le operazioni come codice, limiti gli errori umani e attivi risposte coerenti agli eventi.
- **Applicazione di modifiche frequenti, minime e reversibili**: Progetta i carichi di lavoro per fare in modo che i componenti siano aggiornati regolarmente. Apporta modifiche in incrementi ridotti che possono essere annullati se presentano errori (senza comportare conseguenze per i clienti, ove possibile).
- **Perfeziona frequentemente le procedure operative**: Se usi procedure operative, cerca delle opportunità per migliorarle. Man mano che il tuo carico di lavoro si evolve, garantisci anche l'evoluzione adeguata delle tue procedure. Organizza delle simulazioni regolari per verificare e accertarti che tutte le procedure siano efficaci e che i team le conoscano adeguatamente.

- **Prevedi gli insuccessi**: Esegui un'analisi prefallimentare per individuare le potenziali cause di errore in modo da eliminarle o mitigarle. Testa gli scenari di errore e convalida la tua comprensione relativamente al loro impatto. Testa le tue procedure di risposta per assicurarti che siano efficaci e che i team ne conoscano l'esecuzione. Organizza delle simulazioni regolari per testare i carichi di lavoro e le risposte dei team agli eventi simulati.
- **Impara da tutti gli insuccessi 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.

Definizione

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

- **Organizzazione**
- **Preparazione**
- **Operatività**
- **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 consentire l'integrazione, la distribuzione e la consegna del carico di lavoro consentirà 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. Devi 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. I parametri aziendali e operativi derivati dai risultati aziendali desiderati ti permetteranno di 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

Organizzazione

È necessario che i team abbiano una comprensione condivisa dell'intero carico di lavoro, del ruolo che vi svolgono, 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 i principali stakeholder, 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. Assicurati di essere a conoscenza delle linee guida o degli obblighi definiti dalla governance organizzativa e di fattori esterni, come i requisiti di conformità normativa e gli standard di settore, che possono imporre o sottolineare un'attenzione specifica. Accertati di disporre di meccanismi per identificare le modifiche ai requisiti di governance interna e di conformità esterni. Se non vengono identificati requisiti, assicurati che sia stata applicata la dovuta diligenza per giungere a questa conclusione. Rivedi regolarmente le tue priorità in modo che possano essere aggiornate al mutare delle esigenze.

Valuta le minacce per il business (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 refactoring. 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 devono 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 proprietari di ogni applicazione, carico di lavoro, piattaforma e componente dell'infrastruttura e che per ogni processo e procedura sia identificato un proprietario responsabile della sua definizione e dei proprietari responsabili delle loro 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 ec-

cezioni 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. Gli alti dirigenti sono promotori, sostenitori e motori 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 l'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 conferme parziali. 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 Canone di architettura enfatizza l'apprendimento, la misurazione e il miglioramento. Fornisce una strategia coerente per la valutazione delle architetture e l'implementazione di progetti in grado di ridimensionarsi nel corso del tempo. AWS fornisce lo strumento AWS Well-Architected Tool per aiutarti a rivedere il tuo approccio prima dello sviluppo e lo stato dei tuoi carichi di lavoro prima e durante la fase di produzione. Puoi confrontare il tuo approccio con le best practice architetturali AWS più recenti, monitorare lo stato complessivo dei carichi di lavoro e ottenere informazioni sui potenziali rischi. 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 che possono essere utili per definire le loro priorità.

AWS può aiutarti a istruire i tuoi team su AWS e i suoi servizi, affinché comprendano meglio in che modo le loro scelte possono influire sul carico di lavoro. Per istruire i tuoi team, è consigliabile utilizzare le risorse fornite da AWS Support (AWS Knowledge Center, AWS Discussion Forms e AWS Support Center) e la documentazione AWS.

Se hai domande riguardanti AWS, contatta AWS Support tramite AWS Support Center. AWS condivide inoltre le best practice e i modelli appresi attraverso la gestione di AWS nella Amazon Builders' Library. Un'ampia gamma di altre informazioni utili è disponibile tramite il blog AWS e il podcast ufficiale di AWS. AWS Training and Certification offre risorse di formazione gratuite tramite corsi digitali gestiti dall'utente sulle nozioni di base di AWS. Inoltre, per supportare ulteriormente lo sviluppo delle competenze AWS del tuo team, è possibile iscriversi a corsi di formazione con istruttore.

Per facilitare la gestione dei modelli operativi, è consigliabile utilizzare strumenti o servizi che consentano di gestire centralmente gli ambienti su più account, ad esempio AWS Organizations. Servizi come AWS Control Tower ampliano questa funzionalità di gestione consentendoti di definire piani (a supporto dei tuoi modelli operativi) per configurare gli account, applicare la governance continua tramite AWS Organizations e automatizzare il provisioning di nuovi account. I fornitori di servizi gestiti, come AWS Managed Services, AWS Managed Services Partners o i fornitori di servizi gestiti della 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 a eccellenza operativa . (Per l'elenco completo delle domande e delle best practice relative a eccellenza operativa , consulta l'Appendice.).

OPS 1: In che modo stabilisci quali sono le tue priorità?
È necessario che ognuno capisca 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.
OPS 2: 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 devono 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.
OPS 3: 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.

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. Ripeti le operazioni per sviluppare la telemetria necessaria per monitorare lo stato del carico di lavoro, identificare quando i risultati sono a rischio e abilitare risposte efficaci. Mentre attivi il carico di lavoro, acquisisci un ampio spettro di informazioni per consentire la consapevolezza situazionale (ad esempio cambiamenti di stato, attività utente, accesso con privilegi, contatori di utilizzo), sapendo che hai la possibilità di applicare filtri per selezionare le informazioni più utili nel corso del tempo.

Adotta strategie che migliorino il flusso delle modifiche in produzione e che consentano il refactoring, 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 distribuzione o scoperti negli ambienti.

Adotta prassi che consentano di fornire un feedback rapido sulla qualità e 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 la distribuzione 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. Semplificherai così la risoluzione dei problemi, accelerando la correzione e 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. È consigliabile utilizzare 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 consentirà di trovare eventuali aree che per essere affrontate necessitano di pianificazioni. Predisponi istruzioni che documentano le tue attività di routine e manuali che guidano i processi per la risoluzione dei problemi. Analizza i vantaggi e i rischi per prendere decisioni informate e consentire l'adozione delle modifiche nella produzione.

In AWS, puoi vedere il tuo carico di lavoro completo (applicazioni, infrastruttura, policy, governance e operazioni) in forma di codice. Tutti gli elementi possono essere definiti al suo interno e aggiornati tramite 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. AWS CloudFormation ti consente di avere ambienti di sviluppo, di prova e di produzione sandbox, omogenei e basati su modelli, con livelli crescenti di controllo operativo.

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

OPS 4: In che modo progetti il carico di lavoro al fine di comprenderne lo stato?

Progetta il tuo carico di lavoro in modo da ottenere le informazioni necessarie tra i componenti (ad esempio, parametri, log e tracce) per comprenderne lo stato interno. Ciò ti consente di fornire risposte efficaci in base alle esigenze.

OPS 5: In che modo riduci i difetti, favorisci la correzione e migliori il flusso nella produzione?

Adotta prassi che migliorino il flusso delle modifiche nella produzione, che consentano il refactoring e il feedback veloce su qualità e 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 distribuzione.

OPS 6: In che modo mitighi i rischi della distribuzione?

Adotta prassi che consentano di fornire un feedback rapido sulla qualità e 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 la distribuzione delle modifiche.

OPS 7: Come fai a sapere che sei pronto a supportare un carico di lavoro?

Valuta la disponibilità 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 tag delle

risorse e i Gruppi di risorse AWS 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. 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.

Operatività

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, la distribuzione 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, per capire se stai rispondendo alle esigenze aziendali e per 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 istruzioni precise per gli eventi chiari e ricorri ai manuali 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 trigger 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 panoramiche di pannelli di controllo per i parametri raccolti dai carichi di lavoro e in modo nativo da AWS. Puoi usare CloudWatch o applicazioni di terzi per aggregare e presentare panoramiche al livello di azienda, carico di lavoro e operazioni per le attività operative. AWS offre informazioni sui carichi di lavoro tra-

mite funzionalità di registrazione di log, tra cui AWS X-Ray, CloudWatch, CloudTrail e VPC Flow Logs, che consentono l'identificazione di problemi legati al carico di lavoro per facilitare l'analisi della causa principale e la risoluzione dei problemi.

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

OPS 8: Come fai a comprendere lo stato del tuo carico di lavoro?

Definisci, acquisisci e analizza i parametri del carico di lavoro per ottenere visibilità sugli eventi del carico di lavoro, in modo da intraprendere le azioni appropriate.

OPS 9: 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.

OPS 10: 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.

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

Devi imparare, condividere e migliorare continuamente per sostenere l'eccellenza operativa. Dedica dei cicli di lavoro al raggiungimento di miglioramenti incrementali 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.

In AWS, è possibile esportare i dati di log in Amazon S3 o inviare log direttamente ad Amazon S3 per lo storage a lungo termine. Con AWS Glue, è possibile individuare e preparare i dati di log in Amazon S3 per l'analisi, archiviando i metadati associati nel catalogo dati di AWS Glue. Grazie all'integrazione nativa con Glue, quindi, Ama-

zon Athena può essere utilizzato per analizzare i dati di log, eseguendo query tramite SQL standard. 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.

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

OPS 11: In che modo fai evolvere le operazioni?
Dedica tempo e risorse al miglioramento incrementale 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 a Eccellenza operativa .

Documentazione

- DevOps and AWS

Whitepaper

- Operational Excellence Pillar

Video

- DevOps at Amazon

Sicurezza

Il pilastro Sicurezza include 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 principio di base dell'affidabilità offre una panoramica dei principi di progettazione, delle best practice e delle domande. Puoi trovare una guida prescrittiva sull'implementazione nel whitepaper sul Principio dell'affidabilità.

Principi di progettazione

Esistono sette principi di progettazione per sicurezza nel cloud:

- **Implementa una solida base identitaria**: Implementa il principio del privilegio minore e attua la separazione dei compiti con la corretta autorizzazione per ciascuna interazione con le risorse AWS. Centralizza la gestione delle identità e mira a eliminare la dipendenza dalle credenziali statiche a lungo termine.
- **Abilita la tracciabilità**: Monitora, avvisa e verifica le azioni e le modifiche al tuo ambiente in tempo reale. Integra la raccolta di log e parametri con i sistemi per analizzare e intervenire automaticamente.
- **Applica la sicurezza a tutti i livelli**: Applica un approccio di difesa avanzata con più controlli di sicurezza. Applicalo a tutti i livelli (ad esempio, edge di rete, VPC, bilanciamento del carico, ogni istanza e servizio di elaborazione, sistema operativo, applicazione e codice).
- **Automatizza le best practice per la sicurezza**: I meccanismi di sicurezza automatici basati sul software migliorano la tua capacità di ridimensionare in modo sicuro, più rapido e conveniente. Crea architetture sicure, compresa l'implementazione dei controlli, che sono definite e gestite come codice nei modelli controllati dalle versioni.
- **Proteggi i dati in transito e a riposo**: Classifica i tuoi dati secondo livelli di sensibilità e meccanismi d'uso, come crittografia, tokenizzazione e controllo di accesso, ove opportuno.
- **Tieni le persone a distanza dai 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 di altri errori umani durante la gestione dei dati sensibili.
- **Preparati per gli eventi di sicurezza**: Preparati per un incidente ipotetico creando policy e processi di gestione degli incidenti allineati ai 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 sei aree di best practice per sicurezza nel cloud:

- **Sicurezza**
- **Gestione di identità e accessi (Identity and Access Management)**
- **Rilevamento**

- **Protezione dell'infrastruttura**
- **Protezione dei dati**
- **Risposta agli incidenti**

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 condivisa di AWS permette alle organizzazioni che adottano il cloud di raggiungere i loro obiettivi in termini di sicurezza e conformità. Dato che AWS mette fisicamente in sicurezza l'infrastruttura che supporta i nostri servizi in cloud, come cliente AWS puoi concentrarti sull'utilizzo dei servizi per raggiungere gli obiettivi. AWS Cloud fornisce inoltre l'accesso ai dati sulla sicurezza e offre un approccio automatico per rispondere agli eventi di sicurezza.

Best practice

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 di AWS e del settore nonché con l'intelligence sulle minacce aiuta a sviluppare il modello di rischio e gli obiettivi di controllo. L'automazione dei processi di sicurezza, i test e la convalida consentono di ricalibrare le operazioni di sicurezza.

Le seguenti domande si concentrano su queste considerazioni relative a sicurezza. (Per l'elenco completo delle domande e delle best practice relative a sicurezza, consulta l'Appendice.).

SEC 1: 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 di AWS e del settore nonché con l'intelligence sulle minacce aiuta a sviluppare il modello di rischio e gli obiettivi di controllo. L'automazione dei processi di sicurezza, i test e la convalida consentono di ricalibrare le operazioni di sicurezza.

In AWS, è consigliabile separare i diversi carichi di lavoro per account, in base alla loro funzione e ai requisiti di conformità o di sensibilità dei dati.

Gestione di identità e accessi (Identity and Access Management)

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.

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

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

SEC 2: Come si gestisce l'autenticazione per persone e macchine?

Ci sono due tipi di identità da gestire quando ci si avvicina all'utilizzo di carichi di lavoro AWS sicuri. Comprendere il tipo di identità necessaria per gestire e concedere l'accesso ti aiuta a garantire che le identità corrette abbiano accesso alle risorse giuste nelle condizioni adeguate. Identità umane: amministratori, sviluppatori, operatori e utenti finali necessitano di un'identità per accedere agli ambienti e alle applicazioni AWS. Si tratta di membri dell'organizzazione o utenti esterni con cui collabori e che interagiscono con le tue risorse AWS tramite browser Web, applicazioni client o strumenti a riga di comando interattivi. Identità di macchine: le applicazioni di servizio, gli strumenti operativi e i carichi di lavoro necessitano di un'identità per effettuare richieste ai servizi AWS, ad esempio per leggere i dati. Queste identità includono macchine in esecuzione nell'ambiente AWS, ad esempio istanze Amazon EC2 o funzioni AWS Lambda. Puoi gestire le identità di macchine anche per soggetti esterni che necessitano dell'accesso. Inoltre, possono esistere macchine al di fuori di AWS che hanno bisogno di accedere al tuo ambiente AWS.

SEC 3: Come si gestisce l'autenticazione per persone e macchine?

Gestisci le autorizzazioni per controllare l'accesso alle identità di persone e macchine che richiedono l'accesso ad AWS e al 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 dovrebbe essere concesso utilizzando un approccio con privilegi minimi con le migliori pratiche, inclusi i requisiti di password e l'applicazione del MFA. L'accesso programmatico, comprese le chiamate API ai servizi AWS, deve essere eseguito utilizzando credenziali temporanee e con privilegi limitati come quelle emesse da AWS Security Token Service.

AWS offre risorse che possono aiutarti nella gestione dell'identità e degli accessi. Per apprendere le best practice, esplora 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, una verifica dei controlli relativi ai sistemi di informazioni, per assicurarti che le practice 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.

In AWS, puoi implementare controlli investigativi elaborando log, eventi e monitoraggio che consentono audit, analisi automatizzate e notifiche. I log CloudTrail, le chiamate API AWS e CloudWatch forniscono il monitoraggio di parametri con notifiche, mentre AWS Config fornisce la cronologia delle configurazioni. Amazon GuardDuty è un servizio di rilevazione delle minacce che monitora costantemente possibili comportamenti dannosi o non autorizzati così da proteggere i tuoi account e i tuoi carichi di lavoro su AWS. Sono inoltre disponibili log a livello di servizio, ad esempio puoi utilizzare Amazon Simple Storage Service (Amazon S3) per registrare le richieste di accesso.

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

SEC 4: In che modo individui ed esamini gli eventi di sicurezza?
Acquisisci ed analizza gli eventi a partire da log e parametri 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 lega-

li. 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 in locale.

In AWS, è possibile implementare l'ispezione di pacchetti con stato e senza stato, sia utilizzando tecnologie native di AWS, sia utilizzando prodotti e servizi dei partner disponibili attraverso AWS Marketplace. È necessario utilizzare Amazon Virtual Private Cloud (Amazon VPC) per creare un ambiente privato, protetto e scalabile in cui è possibile definire la propria topologia, inclusi gateway, tabelle di routing e subnet pubbliche e private.

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

SEC 5: Come proteggere 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.
SEC 6: 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 ed interne. Le risorse di calcolo includono istanze EC2, container, funzioni di AWS Lambda, servizi di database, dispositivi IoT e altro.

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 in locale. L'applicazione della protezione dei confini, il monitoraggio dei punti di ingresso e di uscita e la registrazione, il monitoraggio e le notifiche completi sono tutti elementi essenziali per un efficace piano di sicurezza delle informazioni.

I clienti AWS sono in grado di adattare o rafforzare la configurazione di Amazon Elastic Compute Cloud (Amazon EC2), di un container di Amazon EC2 Container Service (Amazon ECS) o di un'istanza AWS Elastic Beanstalk e mantenere questa configurazione su un'Amazon Machine Image (AMI) immutabile. Quindi, che siano attivati da Auto Scaling o lanciati manualmente, tutti i nuovi server virtuali (istanze) lanciati con questa AMI utilizzeranno 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.

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

- Come cliente AWS 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 da AWS o gestita da te.
- È disponibile la registrazione dettagliata che contiene contenuti importanti, come l'accesso ai file e le modifiche.
- AWS ha progettato sistemi di archiviazione con 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 il trasferimento di dati tra regioni. Il contenuto inserito in una Regione rimarrà in quella Regione a meno che tu non abiliti esplicitamente una funzione o utilizzi un servizio che fornisce tale funzionalità.

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

SEC 7: In che modo classificare 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.

SEC 8: In che modo proteggere i dati inattivi?

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

SEC 9: Come proteggere i dati in transito?

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

AWS offre molteplici mezzi per crittografare i dati inattivi 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. È inoltre possibile disporre che l'intero processo di crittografia e decrittografia HTTPS (generalmente noto come terminazione SSL) sia 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 garantire che la tua architettura sia in grado di supportare indagini e ripristini tempestivi.

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

- Sono disponibili registrazioni dettagliate che contengono contenuti importanti, come l'accesso ai file e le modifiche.
- Gli eventi possono essere elaborati automaticamente e possono attivare strumenti che automatizzano le risposte mediante l'uso delle API di AWS.
- Puoi effettuare il pre-provisioning degli strumenti e una "clean room" utilizzando AWS CloudFormation. Questo permette di effettuare indagini forensi in un ambiente sicuro e isolato.

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

SEC 10: In che modo è possibile prevedere gli eventi, rispondervi e risolverli?

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'organizzazione.

Assicurati 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.

Risorse

Consulta le seguenti risorse per ulteriori informazioni sulle best practice relative a Sicurezza.

Documentazione

- AWS Cloud Security
- AWS Compliance
- AWS Security Blog

Whitepaper

- Security Pillar
- AWS Security Overview
- AWS Security Best Practices
- AWS Risk and Compliance

Video

- AWS Security State of the Union
- Shared Responsibility Overview

Affidabilità

Il pilastro Affidabilità include la capacità di un carico di lavoro di svolgere la funzione prevista in modo corretto e coerente quando previsto, ciò include la possibilità di utilizzare e testare il carico di lavoro durante il suo ciclo di vita totale.

Il principio di base dell'affidabilità offre una panoramica dei principi di progettazione, delle best practice e delle domande. Puoi trovare una guida prescrittiva sull'implementazione nel whitepaper sul Principio dell'affidabilità.

Principi di progettazione

Esistono cinque principi di progettazione per affidabilità nel cloud:

- **Ripristino automatico dagli errori**: Monitorando gli indicatori chiave di prestazione (KPI) di un carico di lavoro, è possibile attivare l'automazione in caso di superamento di una soglia. Questi KPI 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 in locale, spesso vengono eseguiti test per dimostrare che il carico di lavoro funziona in uno scenario specifico. I

test non vengono generalmente utilizzati per convalidare le strategie di recupero. Nel cloud, puoi testare il modo in cui il carico di lavoro incorre nell'errore e convalidare le procedure di ripristino. È possibile utilizzare l'automazione per simulare diversi errori o per 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.

- **Dimensionamento 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 su molteplici risorse più piccole per garantire che non condividano un punto di errore comune.
- **Non sarà più necessario indovinare la capacità**: Una causa comune di guasti nei carichi di lavoro in locale è la saturazione delle risorse, quando le richieste assegnate ad 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 ottimale, al fine di soddisfare la domanda senza un provisioning eccessivo o inferiore. Esistono ancora dei limiti, ma alcune quote possono essere controllate e altre possono essere gestite (consulta Gestisci quote e vincoli di servizio).
- **Gestione del cambiamento nell'automazione**: Le modifiche all'infrastruttura dovrebbero essere apportate utilizzando l'automazione. Le modifiche che devono essere gestite includono le modifiche all'automazione, che possono quindi essere monitorate e revisionate.

Definizione

Esistono quattro aree di best practice per affidabilità nel cloud:

- **Fondamenti**
- **Architettura del carico di lavoro**
- **Gestione delle modifiche**
- **Gestione degli errori**

Per ottenere affidabilità, è necessario iniziare dalle basi: un ambiente in cui le quote di servizio e la 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 gli errori. Il carico di lavoro deve gestire le variazioni nella domanda o nei requisiti e deve essere progettato per rilevare l'errore e correggersi automaticamente.

Best practice

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, devono essere stabiliti i requisiti fondamentali che influenzano l'affidabilità. Ad esempio, è necessario disporre di una larghezza di banda di rete sufficiente verso il data center.

Con AWS, la maggior parte di questi requisiti di base è già incorporata o può essere affrontata in base alle esigenze. Il cloud è progettato per essere quasi illimitato, perciò è responsabilità di AWS soddisfare i requisiti di capacità di rete e di elaborazione sufficienti, lasciandoti libero di modificare le dimensioni delle risorse e le allocazioni on demand.

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

REL 1: Come si gestiscono quote e vincoli di servizio?

Per le architetture di carichi di lavoro basate sul cloud, esistono quote di servizio (definite anche come restrizioni dei servizi). Queste quote sono presenti per evitare di effettuare accidentalmente il provisioning di più risorse di quelle necessarie e limitare i tassi di richiesta sulle operazioni API in modo da proteggere i servizi da un uso illecito. Esistono anche vincoli di risorse, ad esempio la velocità con cui è possibile trasferire i bit su un cavo in fibra ottica o la quantità di storage su un disco fisico.

REL 2: Come si pianifica la topologia di rete?

I carichi di lavoro sono spesso presenti in più ambienti. Questi includono più ambienti cloud (sia pubblicamente accessibili 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.

Per le architetture di carichi di lavoro basate sul cloud, esistono quote di servizio (definite anche come restrizioni dei servizi). Queste quote sono presenti per evitare di effettuare accidentalmente il provisioning di più risorse di quelle necessarie e limitare i tassi di richiesta sulle operazioni API in modo da proteggere i servizi da un uso illecito. I carichi di lavoro sono spesso presenti in più ambienti. È necessario monitorare e gestire queste quote per tutti gli ambienti dei carichi di lavoro. Questi includono più ambienti cloud (sia pubblicamente accessibili sia privati) e possono includere 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.

Architettura del carico di lavoro

Un carico di lavoro affidabile comincia con decisioni iniziali di progettazione sia per il software che per l'infrastruttura. Le tue scelte di architettura avranno un impatto sul comportamento del carico di lavoro su tutti e cinque i pilastri di 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. Gli SDK AWS semplificano la scrittura di codici fornendo API specifiche dei linguaggi per i servizi AWS. Questi SDK, oltre alla scelta dei linguaggi, consentono agli sviluppatori di implementare le best practice 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 a affidabilità.

REL 3: Come si progetta l'architettura del servizio di carico di lavoro?
Creazione di 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.
REL 4: 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 in 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 impediscono gli errori e migliorano il tempo medio tra errori (MTBF).
REL 5: 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 consentono ai carichi di lavoro di affrontare stress o guasti, recuperare più rapidamente e mitigare l'impatto di tali problemi. Il risultato è un miglioramento del tempo medio di ripristino (MTTR).

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 in queste reti. I componenti del sistema distribuito devono funzionare in modo da non influire negativamente su altri componenti o sul carico di lavoro.

Gestione delle modifiche

Le modifiche apportate al carico di lavoro o al relativo ambiente devono essere anticipate e sistemate per ottenere un funzionamento affidabile del carico di lavoro. Le modifiche includono quelle imposte al carico di lavoro, ad esempio i picchi di domanda, nonché quelle dall'interno quali le distribuzioni delle caratteristiche e le patch di sicurezza.

Utilizzando AWS, puoi monitorare il comportamento di un carico di lavoro e automatizzare la risposta ai KPI. 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 e controllare la cronologia di tali modifiche.

Le seguenti domande si concentrano su queste considerazioni relative a affidabilità.

REL 6: Come monitorare le risorse del carico di lavoro?
I log e i parametri sono strumenti molto efficaci per ottenere informazioni sullo stato del tuo carico di lavoro. È possibile configurare il carico di lavoro in modo da monitorare i log e i parametri e inviare notifiche quando vengono superate le soglie o si verificano eventi significativi. Il monitoraggio consente al carico di lavoro di riconoscere quando vengono superate le soglie di prestazioni basse o si verificano errori, in modo che possa essere ripristinato automaticamente di rimando.
REL 7: Come si progetta il carico di lavoro per adattarsi ai cambiamenti della domanda?
Un carico di lavoro scalabile fornisce elasticità per aggiungere o rimuovere risorse automaticamente, in modo che vi sia una stretta corrispondenza con la domanda attuale in un dato momento.
REL 8: In che modo implementare le modifiche?
Per distribuire nuove funzionalità e garantire 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 ci si assicura anche che il successo aziendale non diventi un peso. Con il monitoraggio attivo, il tuo team verrà avvisato automaticamente quando gli indicatori KPI si discostano dalle norme previste. La registrazione automatica delle modifiche al proprio ambiente consente di controllare e identificare rapidamente le azioni che potrebbero avere influito sull'affidabilità. I controlli sulla gestione delle modifiche assicurano la possibilità di applicare le regole che garantiscono l'affidabilità di cui hai bisogno.

Gestione degli errori

In qualsiasi sistema di ragionevole complessità è previsto che si verifichino errori. L'affidabilità richiede che il carico di lavoro sia 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 attivare un'azione automatica 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 a affidabilità.

REL 9: In che modo eseguire il backup dei dati?

Esegui il backup dei dati, delle applicazioni e della configurazione per soddisfare i tuoi requisiti relativi agli obiettivi di tempo di ripristino (recovery time objective, RTO) e agli obiettivi di punto di ripristino (recovery point objective, RPO).

REL 10: 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.

REL 11: Come si progetta il carico di lavoro affinché resista ai guasti dei componenti?

I carichi di lavoro con requisiti di disponibilità elevata e MTTR (Mean Time To Recovery) basso devono essere progettati per garantire la resilienza.

REL 12: 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 garantire il funzionamento corretto e offrire la resilienza prevista.

REL 13: Come si pianifica il disaster recovery?

Avere backup e componenti del carico di lavoro ridondanti in loco è l'inizio della strategia di disaster recovery. RTO e RPO sono i tuoi obiettivi per il ripristino della disponibilità. 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.

Esegui regolarmente il backup dei dati e testa i file di backup per assicurarti 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 assicurati che tali test vengano attivati anche dopo importanti cambiamenti del carico di lavoro. Traccia attivamente i KPI, come recovery time objective (RTO) e recovery point objective (RPO), per valutare la resilienza di un carico di lavoro (specialmente

in scenari di test degli errori). Il monitoraggio dei KPI 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 relative a Affidabilità.

Documentazione

- AWS Documentation
- AWS Global Infrastructure
- AWS Auto Scaling: How Scaling Plans Work
- What Is AWS Backup?

Whitepaper

- Reliability Pillar: AWS Well-Architected
- Implementing Microservices on AWS

Efficienza delle prestazioni

Il pilastro Efficienza delle prestazioni include l'abilità di utilizzare in modo efficiente le risorse di elaborazione per soddisfare i requisiti del sistema e conservare tale efficienza a seconda dei cambiamenti della domanda e dell'evoluzione delle tecnologie.

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

Principi di progettazione

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

- **Tecnologie avanzate estese a tutti**: Facilita 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, i da-

tabase NoSQL, la transcodifica multimediale e il machine learning 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.

- **Disponibilità globale in pochi minuti**: Distribuire il carico di lavoro in più regioni AWS in tutto il mondo ti consente di ridurre la latenza e fornire un'esperienza migliore ai tuoi clienti a costi minimi.
- **Utilizzo delle 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 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 con più frequenza**: Le risorse virtuali e automatizzabili ti permettono di portare a termine velocemente i test comparativi utilizzando diversi tipi di istanze, storage e configurazioni.
- **Approccio orientato alla 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 quattro aree di best practice per efficienza delle prestazioni nel cloud:

- **Selezione**
- **Revisione**
- **Monitoraggio**
- **Compromessi**

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.

Rivedendo le tue decisioni a intervalli regolari, avrai la certezza di sfruttare le capacità in continua evoluzione di AWS Cloud. 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, per esempio utilizzando la compressione o la memorizzazione nella cache oppure allentando i requisiti di coerenza.

Best practice

Selezione

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 impiegano funzionalità diverse per migliorare le prestazioni.

Le risorse AWS sono disponibili in numerose tipologie e configurazioni, il che semplifica la ricerca di un approccio che soddisfi appieno le tue esigenze. Inoltre, puoi trovare opzioni che non sono facili da trovare nelle infrastrutture in locale. Ad esempio, un servizio gestito come Amazon DynamoDB offre un database NoSQL interamente gestito, con una latenza di pochissimi millisecondi, indipendentemente dalle dimensioni.

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

PERF 1: In che modo selezioni l'architettura più performante?
Spesso sono necessari molteplici approcci per ottenere prestazioni ottimali in un carico di lavoro. I sistemi Well-Architected utilizzano soluzioni multiple e funzionalità diverse per migliorare le prestazioni.

Quando selezioni i modelli e l'implementazione per la tua architettura, utilizza un approccio basato sui dati per individuare la soluzione ottimale. I solutions architect di AWS, le architetture di riferimento di AWS e i partner AWS Partner Network (APN) possono aiutarti a selezionare un'architettura in base alla conoscenza del settore, ma per ottimizzare la tua architettura saranno necessari i dati ottenuti da benchmark o test di carico.

La tua architettura può riunire vari approcci architetturali (ad esempio basati sugli eventi, ETL o pipeline). L'implementazione della tua architettura sfrutterà i servizi AWS in grado di ottimizzarne le prestazioni. Nelle sezioni seguenti, osserveremo quattro tipi di risorse principali da prendere in considerazione: elaborazione, storage, database e rete.

Calcolo

La selezione delle risorse di calcolo in grado di soddisfare i tuoi requisiti e le tue esigenze di prestazioni e offrire grande efficienza in termini di costi e impegno ti consentirà di ottenere di più con lo stesso numero di risorse. Durante la valutazione delle opzioni di elaborazione, tieni presente i requisiti per le prestazioni del carico di lavoro e i requisiti di costo e utilizzali per prendere decisioni informate.

In AWS, il calcolo avviene in tre forme: istanze, container e funzioni:

- **Le istanze** sono server virtualizzati, che ti consentono di cambiare le loro capacità facendo 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. In AWS, tali istanze di server virtuali sono disponibili in famiglie e dimensioni diverse e offrono un'ampia gamma di funzionalità, tra cui unità a stato solido (SSD) e unità di elaborazione grafica (GPU).
- **I container** rappresentano un metodo di virtualizzazione del sistema operativo con cui puoi eseguire un'applicazione e le relative dipendenze in processi isolati dalle risorse. AWS Fargate è un servizio di elaborazione serverless per container, oppure puoi scegliere Amazon EC2 se hai bisogno di controllare l'installazione, la configurazione e la gestione del tuo ambiente di elaborazione. Puoi anche scegliere tra diverse piattaforme di orchestrazione di 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 ti permette di eseguire un codice senza eseguire un'istanza.

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

PERF 2: In che modo selezioni la soluzione di calcolo?

La soluzione di calcolo ottimale 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 consentire funzioni diverse per migliorare le prestazioni. Selezionare la soluzione di calcolo sbagliata per un'architettura può portare a una riduzione dell'efficienza delle prestazioni.

Quando pianifichi l'utilizzo della capacità di elaborazione, devi sfruttare i meccanismi di elasticità per garantirti una capacità sufficiente a fornire le giuste prestazioni al variare delle esigenze.

Storage

Lo storage sul cloud è un componente fondamentale del cloud computing, poiché predisposto all'archiviazione delle informazioni utilizzate dal carico di lavoro. Lo storage sul cloud è generalmente più affidabile, scalabile e sicuro dei tradizionali sistemi di storage locali. Scegli tra servizi di storage di oggetti, blocchi e file, nonché opzioni di migrazione dei dati nel cloud per il tuo carico di lavoro.

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

- **lo storage di 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 storage di oggetti che offre

scalabilità, disponibilità dei dati, sicurezza e prestazioni leader di 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** fornisce storage a blocchi a disponibilità elevata, costante e a bassa latenza per ogni host virtuale ed è analogo allo storage collegato direttamente (DAS) o a una rete SAN (Storage Area Network). Amazon Elastic Block Storage (Amazon EBS) è stato progettato per carichi di lavoro che richiedono storage persistente accessibile dalle istanze EC2 e consente di ottimizzare le applicazioni con capacità di storage, prestazioni e costi ottimali.
- **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 (EFS) o sono ideali per casi di utilizzo come repository di contenuti di grandi dimensioni, ambienti di sviluppo, store multimediali o home directory. Amazon FSx rende più semplice e conveniente l'avvio e l'esecuzione di file system diffusi in modo da sfruttare le funzionalità avanzate e le prestazioni rapide dei file system open source più utilizzati e con licenza commerciale.

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

PERF 3: In che modo selezioni la soluzione di storage?

La soluzione di storage ottimale per un sistema varia in base a fattori quali: tipo di metodo di accesso (blocco, file od oggetto), schemi di accesso (casuali o sequenziali), throughput necessario, frequenza di accesso (online, offline, archivio), frequenza di aggiornamento (WORM, dinamico) e vincoli di disponibilità e durata. I sistemi Well-Architected utilizzano più soluzioni di storage e consentono funzionalità diverse per migliorare le prestazioni e utilizzare le risorse in modo efficiente.

Nella scelta di una soluzione di storage, accertarsi che sia in linea con gli schemi di accesso sarà cruciale per raggiungere le prestazioni desiderate.

Database

Il cloud offre servizi di database dedicati che risolvono i diversi problemi presentati dal carico di lavoro. Puoi scegliere tra diversi motori di database dedicati, tra cui database relazionali, chiave-valore, documento, in memoria, grafi, serie temporali e libri mastri. Scegliendo il database migliore per risolvere un problema specifico o una serie di problematiche, potrai finalmente abbandonare i database monolitici, restrittivi e indifferenziati e concentrarti sulla creazione di applicazioni in grado di rispondere alle esigenze di prestazioni dei tuoi clienti.

In AWS puoi scegliere tra più motori di database dedicati, tra cui database relazionali, chiave-valore, documento, in memoria, grafi, serie temporali e libri mastri. Con i database AWS, non devi preoccuparti delle attività di gestione del database come provisioning del server, applicazione di patch, installazione, configurazione, backup o ripri-

stino. AWS monitora continuamente i cluster per mantenere i carichi di lavoro operativi con storage con riparazione automatica e auto scaling, in modo da consentirti di concentrarti sullo sviluppo di applicazioni di maggior valore.

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

PERF 4: In che modo selezioni la soluzione di database?

La soluzione di database ottimale per un determinato sistema può variare in base ai requisiti di disponibilità, coerenza, tolleranza della partizione, latenza, durata, scalabilità e capacità di query. Molti sistemi utilizzano diverse soluzioni di database per vari sottosistemi e consentono funzionalità differenti per migliorare le prestazioni. La selezione della soluzione e delle funzionalità errate del database per un sistema può ridurre l'efficienza delle prestazioni.

L'approccio al database del carico di lavoro ha un impatto significativo sull'efficienza delle prestazioni. Spesso è un'area scelta in base alle impostazioni predefinite dell'organizzazione piuttosto che tramite un approccio basato sui dati. E a proposito di storage, è fondamentale prendere in considerazione gli schemi di accesso del tuo carico di lavoro, nonché valutare se altre soluzioni non basate su database potrebbero risolvere il problema in modo più efficiente (ad esempio utilizzare grafici, serie temporali o un database di storage in memoria).

Rete

Poiché la rete si trova tra tutti i componenti del carico di lavoro, può avere grandi ripercussioni positive o negative sulle prestazioni e sul comportamento del carico di lavoro. Esistono anche carichi di lavoro che dipendono in larga misura dalle prestazioni di rete, come nel caso dello High Performance Computing (HPC), dove la comprensione approfondita della rete è importante per migliorare le prestazioni del cluster. È necessario determinare i requisiti del carico di lavoro per larghezza di banda, latenza, jitter e throughput.

In AWS, le reti sono virtualizzate e vengono fornite in una varietà di tipi e configurazioni. Ciò semplifica la scelta delle metodologie di rete più adatte alle tue esigenze. AWS offre caratteristiche dei prodotti (ad esempio reti avanzate, istanze ottimizzate di Amazon EBS, Amazon S3 Transfer Acceleration e Amazon CloudFront dinamico) pensate per l'ottimizzazione del traffico di rete. AWS offre anche funzionalità di rete (ad esempio instradamento della latenza di Amazon Route 53, endpoint VPC di Amazon, AWS Direct Connect e AWS Global Accelerator) per ridurre la distanza di rete o il jitter.

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

PERF 5: In che modo configuri la tua soluzione di rete?

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 in locale, de-

terminano le opzioni di posizione. Questi vincoli possono essere compensati con le edge location o la collocazione delle risorse.

Nella distribuzione della rete devi tener conto dell'ubicazione, puoi decidere di collocare le risorse vicino al punto in cui saranno utilizzate per ridurre la distanza. Utilizza i parametri di rete per apportare modifiche alla configurazione di rete a mano a mano che il carico di lavoro si evolve. Sfruttando elementi quali regioni, gruppi di collocamento e servizi edge, avrai modo di incrementare le prestazioni in maniera significativa. Le reti basate sul cloud possono essere ricostruite o modificate rapidamente, perciò, per mantenere l'efficienza delle prestazioni, l'architettura di rete deve evolvere nel tempo.

Revisione

Le tecnologie cloud sono in rapida evoluzione e devi assicurarti che i componenti del carico di lavoro utilizzino nuove tecnologie e approcci per migliorare continuamente le prestazioni. Devi continuamente valutare e prendere in considerazione le modifiche apportate ai componenti del carico di lavoro per assicurarti di raggiungere gli obiettivi di prestazioni e costi. Le nuove tecnologie, come il machine learning e l'intelligenza artificiale (AI), ti permettono di ridefinire le esperienze dei clienti e di innovare tutti i tuoi carichi di lavoro aziendali.

Sfrutta l'innovazione continua di AWS, orientata alle esigenze dei clienti. Rilasciamo nuove regioni, edge location, servizi e funzionalità a intervalli regolari. Le nuove versioni possono migliorare sensibilmente l'efficienza delle prestazioni della tua architettura.

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

PERF 6: In che modo fai evolvere il carico di lavoro per sfruttare le nuove versioni?

Quando si progettano carichi di lavoro, le opzioni tra cui scegliere sono limitate. Tuttavia, nel tempo diventano disponibili nuove tecnologie e nuovi approcci che potrebbero migliorare le prestazioni.

Le prestazioni scarse delle architetture sono in genere il risultato di un processo di revisione delle prestazioni inesistente o incompleto. Se le prestazioni dell'architettura sono insufficienti, implementare un processo di revisione delle prestazioni ti consentirà di applicare un ciclo PDCA (plan-do-check-act) di Deming per favorire un miglioramento iterativo.

Monitoraggio

Dopo avere implementato il carico di lavoro, è necessario monitorarne le prestazioni in modo da risolvere eventuali problemi prima che influiscano sui clienti. Occorre uti-

lizzare i parametri di monitoraggio per attivare gli allarmi in caso di superamento delle soglie.

Amazon CloudWatch è un servizio di monitoraggio e osservazione che fornisce dati e informazioni utili per monitorare il carico di lavoro, rispondere alle variazioni delle prestazioni a livello di sistema, ottimizzare l'utilizzo delle risorse e ottenere una visione unificata dello stato operativo. CloudWatch raccoglie dati operativi e di monitoraggio sotto forma di log, parametri ed eventi da carichi di lavoro eseguiti su AWS e server in locale. AWS X-Ray aiuta gli sviluppatori ad analizzare ed eseguire il debug della produzione e delle applicazioni distribuite. Con AWS X-Ray, puoi ottenere informazioni approfondite sulle prestazioni dell'applicazione, individuare le cause principali e identificare i colli di bottiglia delle prestazioni. Puoi utilizzare le informazioni ottenute per correggere rapidamente il funzionamento e mantenere le prestazioni del carico di lavoro sempre ottimali.

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

PERF 7: In che modo monitori le tue risorse per assicurarti che abbiano le giuste prestazioni?

Le prestazioni del sistema possono peggiorare nel tempo. Monitora le prestazioni del sistema per identificare l'eventuale riduzione delle prestazioni e rimediare a fattori interni o esterni, come il sistema operativo o il carico dell'applicazione.

Garantire che non vengano visualizzati falsi positivi è fondamentale per una soluzione di monitoraggio efficace. Le attivazioni automatiche prevengono l'errore umano e possono ridurre il tempo necessario per la risoluzione dei problemi. Pianifica giornate di gioco in cui vengono eseguite simulazioni nell'ambiente di produzione, per testare la soluzione di allarme e verificare che riconosca correttamente i problemi.

Compromessi

Quando progetti le soluzioni, pondera i compromessi per garantire una strategia ottimale. A seconda della situazione, puoi accettare dei compromessi in termini di coerenza, durabilità e spazio e favorire il tempo o la latenza allo scopo di garantire prestazioni migliori.

AWS ti consente di raggiungere la disponibilità globale in pochi minuti e distribuire le risorse in più destinazioni nel mondo, al fine di operare a più stretto contatto con gli utenti finali. Inoltre, puoi aggiungere in modo dinamico repliche di sola lettura alle destinazioni di storage delle informazioni, come i sistemi di database, per ridurre il carico sul database principale.

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

PERF 8: Come si utilizzano i compromessi per migliorare le prestazioni?
Quando si progettano soluzioni, determinare i compromessi ti consente di selezionare un approccio ottimale. Spesso è possibile migliorare le prestazioni accettando compromessi in termini di coerenza, durata e spazio a favore di tempo e latenza.

Man mano che apporti modifiche al carico di lavoro, raccogli e valuta i parametri per stabilire l'impatto dei cambiamenti. Misura gli impatti sul sistema e sugli utenti finali per capire in che modo i compromessi adottati influiscono sul carico di lavoro. Adotta un approccio sistematico, come il test del carico, per valutare se i compromessi migliorano le prestazioni.

Risorse

Consulta le seguenti risorse per ulteriori informazioni sulle best practice relative a Efficienza delle prestazioni.

Documentazione

- Amazon S3 Performance Optimization
- Amazon EBS Volume Performance

Whitepaper

- Performance Efficiency Pillar

Video

- AWS re:Invent 2019: Amazon EC2 foundations (CMP211-R2)
- AWS re:Invent 2019: Leadership session: Storage state of the union (STG201-L)
- AWS re:Invent 2019: Leadership session: AWS purpose-built databases (DAT209-L)
- AWS re:Invent 2019: Connectivity to AWS and hybrid AWS network architectures (NET317-R1)
- AWS re:Invent 2019: Powering next-gen Amazon EC2: Deep dive into the Nitro system (CMP303-R2)
- AWS re:Invent 2019: Scaling up to your first 10 million users (ARC211-R)

Ottimizzazione dei costi

Il pilastro Ottimizzazione dei costi include capacità di eseguire sistemi per fornire valore aziendale al minor prezzo possibile

Il principio dell'ottimizzazione dei costi offre una panoramica dei principi di progettazione, delle best practice e delle domande. Puoi trovare una guida prescrittiva sull'implementazione nel whitepaper sul Principio dell'ottimizzazione dei costi.

Principi di progettazione

Esistono cinque principi di progettazione per ottimizzazione dei costi nel cloud:

- **Implementazione della gestione finanziaria nel cloud**: Per migliorare i risultati finanziari e accelerare la realizzazione del valore aziendale nel cloud, devi investire 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 tue funzionalità di sicurezza e operative, devi creare capacità tramite lo sviluppo di competenze, programmi, risorse e processi, per diventare un'organizzazione efficiente in termini di costi.
- **Adozione di un modello di consumo**: Paga solo le risorse di calcolo che richiedi e incrementa o riduci l'utilizzo a seconda dei requisiti aziendali, e non attraverso il ricorso a una previsione elaborata. Ad esempio, gli ambienti di prova e di sviluppo sono generalmente 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 il risultato aziendale del carico di lavoro e i costi associati alla sua produzione. Usa questi dati per conoscere i ricavi che ottieni grazie all'aumento della produttività e alla riduzione dei costi.
- **Smetti di spendere denaro per attività onerose indifferenziate**: AWS si occupa delle attività onerose dei data center come il racking, lo stacking e l'alimentazione dei server. Inoltre, elimina l'onere operativo della gestione di sistemi operativi e applicazioni con servizi gestiti. In questo modo, potrai concentrarti sui tuoi clienti e sui progetti aziendali anziché sull'infrastruttura IT.
- **Analisi e attribuzione della spesa**: Il cloud ti aiuta a individuare con facilità e precisione l'utilizzo e il costo dei sistemi, il che consente quindi l'attribuzione trasparente dei costi IT per i singoli proprietari del carico di lavoro. Questo ti aiuta a misurare il ritorno sull'investimento (ROI) e offre ai proprietari del carico di lavoro la possibilità di ottimizzare le proprie risorse e ridurre i costi.

Definizione

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

- **Esercizio della gestione finanziaria del cloud**
- **Consapevolezza delle spese e dell'utilizzo**

- **Risorse convenienti**
- **Gestione delle risorse di domanda e offerta**
- **Ottimizzazione nel tempo**

Come per gli altri principi di base all'interno del Canone di architettura, occorre considerare alcuni compromessi; ad esempio, è meglio ottimizzare la velocità di commercializzazione o i costi? In alcuni casi, è meglio ottimizzare la velocità: entrare nel mercato rapidamente, distribuire nuove caratteristiche o semplicemente rispettare una scadenza piuttosto che investire nell'ottimizzazione anticipata dei costi. 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 distribuzione più conveniente. Questo potrebbe portare a distribuzioni sovra-assegnate e sotto-ottimizzate. Tuttavia, si tratta di una scelta ragionevole quando devi trasferire le risorse dal tuo ambiente locale al cloud ed eseguire l'ottimizzazione di conseguenza. Investire in anticipo la giusta quantità di energia in una strategia di ottimizzazione dei costi consente di realizzare i vantaggi economici del cloud in modo più rapido, assicurando il rispetto costante delle best practice ed evitando un provisioning superfluo. 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

Esercizio della gestione finanziaria del cloud

Con l'adozione del cloud, i team tecnologici innovano più rapidamente grazie a cicli di approvazione, approvvigionamento e distribuzione 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.

In AWS puoi utilizzare Cost Explorer e, facoltativamente, Amazon Athena e Amazon QuickSight, con il report costi e utilizzo (CUR) per fornire consapevolezza su costi e utilizzo in tutta l'organizzazione. Budget AWS fornisce notifiche proattive relative a costi e utilizzo. I blog AWS forniscono informazioni su nuovi servizi e caratteristiche per consentirti di essere sempre aggiornato sulle nuove versioni dei servizi.

Le seguenti domande si concentrano su queste considerazioni relative a ottimizzazione dei costi. (Per l'elenco completo delle domande e delle best practice relative a ottimizzazione dei costi, consulta l'Appendice.).

COST 1: Come implementi la gestione finanziaria nel cloud?

L'implementazione della gestione finanziaria del cloud consente alle organizzazioni di conseguire un valore aggiunto e il successo finanziario ottimizzando i costi e l'utilizzo e ricalibrando le risorse in AWS.

Quando crei una funzione di ottimizzazione dei costi, puoi utilizzare i membri e integrare il team con esperti di gestione finanziaria del cloud e ottimizzazione 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.

Consapevolezza delle spese e dell'utilizzo

La maggiore flessibilità e agilità consentite dal cloud incoraggiano l'innovazione, lo sviluppo e la distribuzione rapidi. Elimina i processi manuali e il tempo associati al provisioning dell'infrastruttura locale, 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.

Con AWS puoi creare una struttura di account con AWS Organizations o AWS Control Tower per garantire la separazione e semplificare l'allocazione dii costi e utilizzo. Puoi anche utilizzare il tagging delle risorse per associare informazioni aziendali e organizzative a utilizzo e costi. Utilizza AWS Cost Explorer per osservare costi e utilizzo, oppure crea analisi e pannelli di controllo personalizzati con Amazon Athena e Amazon QuickSight. Puoi verificare costi e utilizzo con le notifiche di Budget AWS e controllarli usando AWS Identity and Access Management (IAM) e quote di servizio.

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

COST 2: In che modo gestisci l'utilizzo?
Stabilisci policy e meccanismi per assicurarti di sostenere costi adeguati mentre raggiungi gli obiettivi. Utilizzando un approccio di controllo e bilanciamento reciproco, è possibile innovare senza spendere troppo.
COST 3: In che modo monitori l'utilizzo e il costo?
Stabilisci policy e procedure per monitorare e allocare i costi in modo appropriato. Ciò ti consente di misurare e migliorare l'efficienza in termini di costi del carico di lavoro.
COST 4: In che modo disattivi le risorse?
Implementa il controllo del cambiamento e la gestione delle risorse dall'inizio del progetto alla fine del ciclo di vita. In questo modo, puoi chiudere o interrompere le risorse non utilizzate 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 dei tag alle tue risorse AWS (come le istanze EC2 o i bucket S3), AWS genera un report su costi e utilizzo con i tuoi tag e i dati sul tuo utilizzo. 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.

Assicurati di utilizzare il giusto livello di dettaglio e granularità quando crei report e monitori costi e utilizzo. Per un approfondimento di alto livello di informazioni e tendenze, utilizza la granularità giornaliera di AWS Cost Explorer. Per analisi e ispezioni più specifiche, utilizza la granularità oraria di AWS Cost Explorer o Amazon Athena e Amazon QuickSight con il report costi e utilizzo a 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

Utilizzare risorse e istanze adeguate al tuo carico di lavoro è fondamentale per ridurre i costi. Ad esempio, un processo di reporting 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 basato sul Canone di architettura AWS si basa sulle 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 un'ampia gamma di offerte flessibili e convenienti per acquisire istanze da EC2 e altri servizi per soddisfare al meglio le tue necessità. Le *istanze on demand* ti consentono di pagare la capacità di elaborazione a ore e non richiedono impegni minimi. *Savings Plans e le istanze riservate* offrono risparmi fino al 75% rispetto ai prezzi on demand. Con le istanze Spot, puoi sfruttare la capacità inutilizzata di Amazon EC2 e risparmiare fino al 90% sui prezzi on demand. Le *istanze Spot* risultano adeguate quando il sistema può tollerare l'utilizzo di un parco server in cui i singoli server possano andare e venire dinamicamente, come server web stateless, elaborazioni batch o quando si usano HPC e Big Data.

Anche la scelta del servizio appropriato può ridurre l'utilizzo e i costi; ad esempio, CloudFront può ridurre al minimo il trasferimento dei dati o eliminare del tutto i costi, mentre l'utilizzo di Amazon Aurora su RDS può rimuovere gli elevati costi di licenza dei database.

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

COST 5: In che modo valuti i costi quando selezioni i servizi?

Amazon EC2, Amazon EBS e Amazon S3 sono servizi AWS del blocco predefinito. I servizi gestiti, come Amazon RDS e Amazon DynamoDB, sono servizi AWS di livello superiore o di livello applicazione. 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.

COST 6: 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.

COST 7: 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.

COST 8: 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.

Scomponendo i costi durante la selezione del servizio e usando strumenti come Cost Explorer e AWS Trusted Advisor per esaminare con regolarità l'uso di AWS, puoi attivamente monitorare il tuo utilizzo e modificare le distribuzioni di conseguenza.

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, eliminando così la

necessità di un provisioning superfluo costoso e dispendioso. Puoi anche gestire la domanda utilizzando tecniche come throttling, 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 puoi predisporre automaticamente le risorse da associare alla domanda di carico di lavoro. Auto Scaling con strategie basate su domanda o tempo ti consente di aggiungere e rimuovere le risorse in base alle esigenze. Se riesci a prevedere le variazioni nella domanda, puoi risparmiare di più e assicurarti che le risorse corrispondano alle esigenze del tuo carico di lavoro. Puoi usare Amazon API Gateway per implementare il throttling o Amazon SQS per implementare una coda nel tuo carico di lavoro. Entrambi consentono di modificare la richiesta nei componenti del carico di lavoro.

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

COST 9: Come gestisci la domanda e fornisci le risorse?

Per avere un carico di lavoro con costo e prestazioni bilanciate, assicurati che venga utilizzato tutto ciò per cui paghi ed evita le istanze molto sottoutilizzate. Un parametro di utilizzo distorto, in qualsiasi delle suddette direzioni, ha un impatto negativo sull'organizzazione, sia per i costi operativi (basse prestazioni a causa di un utilizzo eccessivo) che per le spese AWS sprecate (a causa di un provisioning eccessivo).

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, assicurati 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

Poiché AWS rilascia nuovi servizi e caratteristiche, è consigliabile rivedere le decisioni correnti sull'architettura per garantire 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 caratteristiche 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 essere sicuro di 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 a ottimizzazione dei costi.

COST 10: In che modo valuti i nuovi servizi?

Poiché AWS rilascia nuovi servizi e funzionalità, è consigliabile rivedere le decisioni correnti sull'architettura per garantire che continuino a essere le più convenienti.

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

Risorse

Consulta le seguenti risorse per ulteriori informazioni sulle best practice relative a Ottimizzazione dei costi.

Documentazione

- AWS Documentation

Whitepaper

- Cost Optimization Pillar

Il processo di revisione

La revisione delle architetture deve essere 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) che è una conversazione piuttosto che 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 Canone di architettura per eseguire costantemente la revisione della loro architettura, piuttosto che fare una riunione di revisione formale. Un approccio 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.

AWS Well-Architected è allineato alla modalità interna di revisione dei sistemi e dei servizi di AWS. Si basa su un insieme di principi di progettazione che influenzano l'approccio architetturale e su domande che garantiscano che le persone non trascurino aree che spesso figurano nell'Analisi della causa principale (RCA). Ogni volta che si presenta un problema significativo con un sistema interno, un servizio AWS o un cliente, ci serviamo dell'RCA per vedere se possiamo migliorare il processo di revisione utilizzato.

Le revisioni devono essere applicate a tappe fondamentali nel ciclo di vita del prodotto, all'inizio della fase di progettazione per evitare *decisioni unidirezionali* [1] che sono difficili da modificare prima della data di implementazione. Dopo l'implementazione il carico di lavoro continuerà ad evolversi man mano che aggiungi funzionalità e modifichi le implementazioni della tecnologia. 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 o misura indipendente dovrai assicurarti che nella conversazione siano presenti 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

[1] Molte decisioni sono reversibili e quindi bidirezionali. Queste decisioni possono usare un processo leggero. Le decisioni unidirezionali sono difficili o impossibile da annullare e richiedono un'ispezione maggiore prima di essere prese.

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
- Lista di azioni delle domande che richiedono ricerche fuori banda per le risposte (ad esempio, «abbiamo abilitato la crittografia o no?»)

Dopo aver eseguito la revisione dovresti avere un elenco di problemi a cui assegnare delle priorità sulla base del contesto aziendale. Dovrai anche prendere in considerazione l'impatto di tali problemi sul lavoro quotidiano del tuo team. Se affronti questi problemi in anticipo puoi liberare del tempo per lavorare sulla creazione di valore aziendale piuttosto che risolvere i problemi ricorrenti. Man mano che affronti i problemi puoi aggiornare la revisione per vedere se 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 aiuta a comprendere 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 utilizzare 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
- «We don’t want others to know the secrets of our solution implementation!»
  - Se poni le domande del Canone di architettura, 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 Canone di architettura AWS fornisce best practice architetturali relative a cinque pilastri per la progettazione e la gestione di sistemi affidabili, sicuri, efficienti e a costi contenuti nel cloud. Il canone fornisce un insieme di domande che ti permettono di eseguire la revisione di un'architettura esistente o proposta. Il canone fornisce un insieme di best practice AWS per ogni pilastro. L'utilizzo del canone nella tua architettura ti aiuta a produrre sistemi stabili ed efficienti, che ti permettono di concentrarti sui tuoi requisiti funzionali.

Collaboratori

Hanno contribuito alla stesura di questo documento:

- Rodney Lester: Senior Manager Well-Architected, Amazon Web Services
- Brian Carlson: Operations Lead Well-Architected, Amazon Web Services
- Ben Potter: Security Lead Well-Architected, Amazon Web Services
- Eric Pullen: Performance Lead Well-Architected, Amazon Web Services
- Seth Eliot: Reliability Lead Well-Architected, Amazon Web Services
- Nathan Besh: Cost Lead Well-Architected, Amazon Web Services
- Jon Steele: Senior Technical Account Manager, Amazon Web Services
- Ryan King: Technical Program Manager, Amazon Web Services
- Erin Rifkin: Senior Product Manager, Amazon Web Services
- Max Ramsay: Principal Security Solutions Architect, Amazon Web Services
- Scott Paddock: Security Solutions Architect, Amazon Web Services
- Callum Hughes: Solutions Architect, Amazon Web Services

Approfondimenti

AWS Cloud Compliance

AWS Well-Architected Partner program

AWS Well-Architected Tool

AWS Well-Architected homepage

Cost Optimization Pillar whitepaper

Operational Excellence Pillar whitepaper

Performance Efficiency Pillar whitepaper

Reliability Pillar whitepaper

Security Pillar whitepaper

The Amazon Builders' Library

Revisioni del documento

Tabella 2. Revisioni principali:

Data	Descrizione
Luglio 2020	Revisione e riscrittura di molte domande e risposte.
Luglio 2019	Aggiunta di AWS Well-Architected Tool, collegamenti a AWS Well-Architected Labs e AWS Well-Architected Partners, correzioni minori per abilitare la versione in più lingue del canone.
Novembre 2018	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, componenti, ecc.). Presentazione delle domande modificata per includere il testo descrittivo.
Giugno 2018	Aggiornamenti volti a semplificare il testo delle domande e a migliorare la leggibilità.
Novembre 2017	Eccellenza operativa spostata all'inizio della sezione sui pilastri e riscritta in modo che inquadri gli altri pilastri. Aggiornamenti di altri pilastri per riflettere l'evoluzione di AWS.
Novembre 2016	Framework aggiornato per includere i pilastri dell'eccellenza operativa; altri pilastri rivisti e aggiornati per ridurre la duplicazione e incorporare le nozioni apprese grazie alle revisioni eseguite con migliaia di clienti.
Novembre 2015	Appendice aggiornata con le informazioni attuali su Amazon CloudWatch Logs.
Ottobre 2015	Pubblicazione originale.

Appendice: domande e best practice

Eccellenza operativa

Organizzazione

OPS 1 In che modo stabilisci quali sono le tue priorità?

È necessario che ognuno capisca 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:

- **Valutazione delle esigenze dei clienti esterni**: Coinvolgi i principali stakeholder, compresi i team aziendali, di sviluppo e operativi, per determinare dove concentrare gli sforzi in base alle esigenze dei clienti esterni. Questo ti garantirà una conoscenza approfondita del supporto operativo necessario per raggiungere i risultati aziendali desiderati.
- **Valutazione delle esigenze dei clienti interni**: Coinvolgi i principali stakeholder, 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.
- **Valutazione dei requisiti di governance**: Assicurati di conoscere le linee guida o gli obblighi definiti dalla tua organizzazione che possono imporre o enfatizzare l'attenzione ad aspetti specifici. Valuta i fattori interni, come policy, standard e requisiti dell'organizzazione. Accertati di disporre di meccanismi per identificare le modifiche alla governance. Se non vengono identificati requisiti di governance, assicurati che sia stata applicata la dovuta diligenza per giungere a questa conclusione.
- **Valutazione dei requisiti di conformità**: Valuta i fattori esterni, come i requisiti di conformità normativa e gli standard di settore, per assicurarti di essere a conoscenza delle linee guida o degli obblighi che possono imporre o sottolineare un'attenzione specifica. Se non vengono identificati requisiti di conformità, assicurati di applicare la dovuta diligenza a questa determinazione.
- **Valutazione del 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.
- **Valutazione dei compromessi**: 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.

- **Gestione dei vantaggi e dei rischi**: Gestisci i vantaggi e i rischi per prendere decisioni informate nel determinare dove concentrare gli sforzi. Ad esempio, può essere vantaggioso distribuire un sistema con problemi irrisolti, in modo da mettere a disposizione dei clienti nuove funzionalità importanti. Può essere possibile ridurre i rischi associati o la presenza di un rischio potrebbe diventare inaccettabile, nel qual caso si intraprenderà un'azione per risolverlo.

OPS 2 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 devono 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:

- **Per le risorse esistono proprietari identificati**: È utile comprendere chi ha la proprietà di ogni applicazione, carico di lavoro, piattaforma e componente dell'infrastruttura, qual è il valore aziendale fornito da tale componente e perché tale proprietà esiste. Comprendere il valore aziendale di questi singoli componenti e il modo in cui supportano i risultati aziendali fornisce indicazioni sui processi e le procedure applicati.
- **Per processi e procedure esistono proprietari identificati**: È utile comprendere chi ha la proprietà della definizione di singoli processi e procedure, perché tali processi e procedure specifici vengono utilizzati e perché tale proprietà esiste. Comprendere i motivi per cui vengono utilizzati processi e procedure specifici consente di identificare le opportunità di miglioramento.
- **Per le attività operative esistono proprietari identificati responsabili delle loro prestazioni**: È utile comprendere chi ha la responsabilità di eseguire attività specifiche su carichi di lavoro definiti e perché tale responsabilità esiste. Comprendere chi ha la responsabilità di eseguire le attività fornisce indicazioni su eseguirà l'attività, su chi convaliderà il risultato e su chi fornirà feedback al proprietario dell'attività.
- **I membri del team sanno di cosa sono responsabili**: Comprendere le responsabilità del tuo ruolo e il modo in cui contribuisci ai risultati aziendali fornisce indicazioni sulle priorità delle tue attività e sul perché il tuo ruolo è importante. In questo modo i membri del team possono riconoscere le esigenze e rispondere in modo appropriato.
- **Esistono meccanismi per identificare responsabilità e proprietà**: Quando non viene identificato alcun individuo o team, esistono percorsi di escalation definiti nei confronti di soggetti dotati dell'autorità per assegnare la proprietà o la pianificazione connesse al soddisfacimento dell'esigenza in questione.
- **Esistono meccanismi per richiedere aggiunte, modifiche ed eccezioni**: È possibile effettuare richieste ai proprietari di processi, procedure e risorse. Prendi decisioni informate per approvare le richieste quando vengono ritenute fattibili e appropriate dopo una valutazione dei vantaggi e dei rischi.
- **Le responsabilità tra i team sono predefinite o negoziate**: Esistono accordi definiti o negoziati tra i team che descrivono come funzionano e si supportano reciprocamente (ad

esempio, tempi di risposta, obiettivi o contratti relativi al livello di servizio). 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.

OPS 3 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:

- **Sponsorizzazione esecutiva**: Gli alti dirigenti stabiliscono chiaramente le aspettative per l'organizzazione e valutano il successo. Gli alti dirigenti sono promotori, sostenitori e motori per l'adozione delle best practice e l'evoluzione dell'organizzazione.
- **Intervento dei membri del team quando i risultati sono a rischio**: Il proprietario del carico di lavoro definisce le linee guida e l'ambito consentendo ai membri del team di rispondere quando i risultati sono a rischio. I meccanismi di escalation vengono utilizzati ai fini dell'orientamento quando gli eventi sono al di fuori dell'ambito definito.
- **Incoraggiamento all'escalation**: I membri del team dispongono di meccanismi e sono incoraggiati a segnalare le preoccupazioni ai responsabili delle decisioni e agli stakeholder se ritengono che i risultati sono a rischio. L'escalation deve essere eseguita in anticipo e di frequente, in modo che i rischi possano essere identificati e limitati prima che provochino incidenti.
- **Comunicazioni tempestive, chiare e fruibili**: Esistono meccanismi che vengono utilizzati per fornire tempestivamente notifiche ai membri del team in merito a rischi noti ed eventi pianificati. Laddove è possibile, vengono forniti contesto, dettagli e tempo per determinare se è necessario intervenire, in che modo e con quali tempistiche. Ad esempio, si può essere emettere un avviso di vulnerabilità del software in modo che le patch vengano applicate rapidamente, oppure si può fornire un avviso sulle promozioni di vendita pianificate al fine di bloccare le modifiche per evitare il rischio di interruzione del servizio.
- **Incoraggiamento alla sperimentazione**: La sperimentazione accelera l'apprendimento e mantiene acceso l'interesse e il coinvolgimento dei membri del team. Un risultato indesiderato è un esperimento riuscito tramite il quale viene identificato un percorso che non porterà al successo. I membri del team non vengono puniti per gli esperimenti riusciti con risultati indesiderati. La sperimentazione è necessaria per realizzare l'innovazione e trasformare le idee in risultati.
- **Autorizzazione e incoraggiamento ai membri del team a mantenere e ampliare le proprie competenze**: I team devono aumentare le proprie competenze per adottare nuove tecnologie e supportare i cambiamenti 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 l'apprendimento.

- **Risorse appropriate per i team**: Mantieni la capacità dei membri del team e fornisci strumenti e risorse per supportare le esigenze del carico di lavoro. I membri del team con troppe mansioni aumentano il rischio di incidenti causati da errori umani. Gli investimenti in strumenti e risorse (ad esempio, fornendo automazione per le attività eseguite di frequente) possono ricalibrare l'efficacia del team, consentendogli di supportare attività aggiuntive.
- **Incoraggiamento e ricerca di opinioni diverse all'interno e tra i team**: 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 conferme parziali. Aumenta l'inclusione, la diversità e l'accessibilità all'interno dei team per ottenere prospettive vantaggiose.

Preparazione

OPS 4 In che modo progetti il carico di lavoro al fine di comprenderne lo stato?

Progetta il tuo carico di lavoro in modo da ottenere le informazioni necessarie tra i componenti (ad esempio, parametri, log e tracce) per comprenderne lo stato interno. Ciò ti consente di fornire risposte efficaci in base alle esigenze.

Best practice:

- **Implementazione della telemetria dell'applicazione**: Implementa il codice dell'applicazione affinché fornisca informazioni sullo stato interno, sullo stato e sul raggiungimento dei risultati aziendali, ad esempio lunghezza della coda, messaggi di errore e tempi di risposta. Utilizza queste informazioni per stabilire quando è necessaria una risposta.
- **Implementazione e configurazione della telemetria del carico di lavoro**: Progetta e configura il carico di lavoro affinché fornisca informazioni sul suo stato interno e sullo stato corrente, ad esempio, volume delle chiamate API, codici di stato HTTP ed eventi di scalabilità. Utilizza queste informazioni per determinare quando è necessaria una risposta.
- **Implementazione della telemetria dell'attività degli utenti**: Implementa il codice dell'applicazione affinché fornisca informazioni sulle attività degli utenti, ad esempio, flussi di clic o transazioni avviate, abbandonate e completate. Utilizza queste informazioni per comprendere come viene utilizzata l'applicazione, i modelli di utilizzo e per stabilire quando è necessaria una risposta.
- **Implementazione della telemetria delle dipendenze**: Progetta e configura il carico di lavoro affinché fornisca informazioni sullo stato (ad esempio, raggiungibilità o tempo di risposta) delle risorse da cui dipende. Esempi di dipendenze esterne possono includere database esterni, DNS e connettività di rete. Utilizza queste informazioni per stabilire quando è necessaria una risposta.
- **Implementazione della tracciabilità delle transazioni**: Implementa il codice dell'applicazione e configura i componenti del carico di lavoro affinché forniscano informazioni sul flusso delle transazioni nel carico di lavoro. Utilizza queste informazioni per stabilire quando è necessaria una risposta e per favorire l'identificazione dei fattori che contribuiscono all'origine di un problema.

OPS 5 In che modo riduci i difetti, favorisci la correzione e migliori il flusso nella produzione?

Adotta prassi che migliorino il flusso delle modifiche nella produzione, che consentano il refactoring e il feedback veloce su qualità e 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 distribuzione.

Best practice:

- **Utilizzo del controllo delle versioni**: Utilizza il controllo delle versioni per abilitare il monitoraggio di modifiche e rilasci.
- **Test e convalida delle modifiche**: Testa e convalida le modifiche per limitare e rilevare gli errori. Automatizza il testing per ridurre gli errori causati dai processi manuali e il livello di impegno richiesto per il test.
- **Utilizzo di sistemi di gestione delle configurazioni**: Utilizza sistemi di gestione delle configurazioni per apportare modifiche alla configurazione e tenerne traccia. Questi sistemi riducono gli errori causati dai processi manuali e il livello di impegno richiesto per la distribuzione delle modifiche.
- **Utilizzo di sistemi di gestione della creazione e distribuzione**: Utilizza sistemi di gestione della creazione e distribuzione Questi sistemi riducono gli errori causati dai processi manuali e il livello di impegno richiesto per la distribuzione delle modifiche.
- **Esecuzione della 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 e il livello di impegno richiesto per applicare le patch.
- **Condivisione degli standard di progettazione**: Condividi le best practice con i team per incrementare la consapevolezza e potenziare al massimo i vantaggi delle attività di sviluppo.
- **Implementazione di prassi per migliorare la qualità del codice**: Implementa prassi per migliorare la qualità del codice e ridurre al minimo i difetti, ad esempio sviluppo basato su test, revisioni del codice e adozione di standard.
- **Utilizzo di più ambienti**: Utilizza ambienti multipli per sperimentare, sviluppare e testare il carico di lavoro. Utilizza livelli crescenti di controlli man mano che gli ambienti si avvicinano alla fase di produzione per avere la certezza che il carico di lavoro funzionerà come previsto una volta distribuito.
- **Applicazione di modifiche frequenti, minime e reversibili**: Le modifiche frequenti, minime e reversibili riducono la portata e l'impatto di una modifica. Questo semplifica la risoluzione dei problemi, consente tempi di correzione più rapidi e permette di eseguire il rollback di una modifica.
- **Automazione completa dell'integrazione e della distribuzione**: Automatizza la creazione, la distribuzione e il test del carico di lavoro. Questo riduce gli errori causati dai processi manuali e l'impegno necessario per distribuire le modifiche.

OPS 6 In che modo mitighi i rischi della distribuzione?

Adotta prassi che consentano di fornire un feedback rapido sulla qualità e 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 la distribuzione delle modifiche.

Best practice:

- **Preparazione di un piano in caso di esito negativo delle modifiche**: Pianifica il ripristino di uno stato corretto noto o la correzione nell'ambiente di produzione nel caso in cui una modifica non produca il risultato desiderato. Questa preparazione riduce i tempi di ripristino grazie a risposte più veloci.
- **Test e convalida delle modifiche**: Testa le modifiche e convalida i risultati in tutte le fasi del ciclo di vita per confermare le nuove funzionalità e ridurre al minimo il rischio e l'impatto delle distribuzioni non riuscite.
- **Utilizzo di sistemi di gestione della distribuzione**: Usa sistemi di gestione della distribuzione per monitorare e implementare una modifica. Questo riduce gli errori causati dai processi manuali e l'impegno necessario per distribuire le modifiche.
- **Test utilizzando distribuzioni limitate**: Esegui test con distribuzioni limitate accanto ai sistemi esistenti per confermare i risultati desiderati prima della distribuzione su vasta scala. Ad esempio, utilizza test della distribuzione di tipo canary oppure distribuzioni one-box.
- **Distribuzione utilizzando ambienti paralleli**: Implementa le modifiche in ambienti paralleli, quindi esegui la transizione al nuovo ambiente. Mantieni l'ambiente precedente finché non viene confermata la riuscita della distribuzione. In questo modo si riducono i tempi di ripristino grazie alla possibilità di eseguire il rollback all'ambiente precedente.
- **Distribuzione di modifiche frequenti, minime e reversibili**: Utilizza modifiche frequenti, minime e reversibili per ridurre la portata e l'impatto di una modifica. Semplificherai così la risoluzione dei problemi, accelerando la correzione e mantenendo la possibilità di rollback delle modifiche.
- **Automazione completa dell'integrazione e della distribuzione**: Automatizza la creazione, la distribuzione e il test del carico di lavoro. Questo riduce gli errori causati dai processi manuali e l'impegno necessario per distribuire le modifiche.
- **Automazione dei test e del rollback**: Automatizza i test degli ambienti distribuiti per confermare i risultati 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.

OPS 7 Come fai a sapere che sei pronto a supportare un carico di lavoro?

Valuta la disponibilità 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:

- **Verifica della capacità del personale**: Predisponi un meccanismo per stabilire se disponi del numero appropriato di risorse qualificate per supportare le esigenze operative. Forma

il personale e adegua la dotazione di personale, se necessario, per mantenere un supporto efficace.

- **Revisione costante della prontezza operativa**: Assicurati di effettuare una revisione costante della capacità di gestire un carico di lavoro. La revisione deve includere come minimo la prontezza operativa dei team e del carico di lavoro, nonché i requisiti per la sicurezza. Implementa le attività di revisione nel codice e attiva revisioni automatizzate in risposta agli eventi, se del caso, per assicurare coerenza e velocità di esecuzione e per ridurre gli errori causati dai processi manuali.
- **Utilizzo di runbook per eseguire le procedure**: I runbook sono procedure documentate per raggiungere determinati risultati. Abilita risposte coerenti e tempestive a eventi noti documentando le procedure nei runbook. Implementa i runbook come codice e attiva l'esecuzione dei runbook in risposta agli eventi, se del caso, per assicurare coerenza e velocità di risposta e per ridurre gli errori causati dai processi manuali.
- **Utilizzo dei playbook per analizzare i problemi**: Abilita risposte coerenti e tempestive a problemi poco chiari, documentando il processo di verifica 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.
- **Adozione di decisioni informate per distribuire sistemi e modifiche**: Valuta la capacità del team di supportare il carico di lavoro e la conformità del carico di lavoro alla governance. Confronta questi aspetti con i vantaggi della distribuzione quando decidi se eseguire il passaggio di un sistema o di una modifica in produzione. Per prendere decisioni informate, tieni conto dei rischi e dei benefici.

Operatività

OPS 8 Come fai a comprendere lo stato del tuo carico di lavoro?

Definisci, acquisisci e analizza i parametri del carico di lavoro per ottenere visibilità sugli eventi del carico di lavoro, in modo da intraprendere le azioni appropriate.

Best practice:

- **Identificazione degli indicatori chiave delle prestazioni**: Identifica gli indicatori chiave delle prestazioni (KPI) in base ai risultati aziendali desiderati (ad esempio, tasso di ordini, tasso di conservazione dei clienti e profitti rispetto alle spese operative) e ai risultati dei clienti (ad esempio, soddisfazione dei clienti). Valuta i KPI per determinare il successo del carico di lavoro.
- **Definizione dei parametri del carico di lavoro**: Definisci i parametri del carico di lavoro per misurare il raggiungimento dei KPI (ad esempio, carrelli degli acquisti abbandonati, ordini effettuati, costo, prezzo e spesa allocata per il carico di lavoro). Definisci i parametri del carico di lavoro per misurarne lo stato (ad esempio, tempo di risposta dell'interfaccia, percentuale di errori, richieste effettuate, richieste completate e utilizzo). Valuta i parametri per stabilire se il carico di lavoro raggiunge i risultati previsti e per comprendere lo stato del carico di lavoro.

- **Raccolta e analisi dei parametri del carico di lavoro**: Esegui revisioni proattive regolari dei parametri per identificare le tendenze e stabilire dove sono necessarie risposte adeguate.
- **Definizione di baseline per i parametri del carico di lavoro**: Definisci le baseline per i parametri in modo da fornire i valori previsti di base per il confronto e l'identificazione dei componenti con prestazioni basse o alte. Identifica le soglie di miglioramento, verifica e intervento.
- **Acquisizione dei modelli di attività previsti per il carico di lavoro**: Definisci modelli di attività del carico di lavoro per identificare comportamenti anomali in modo da rispondere in modo appropriato, se necessario.
- **Attivazione di un avviso quando i risultati del carico di lavoro sono a rischio**: Attiva un avviso quando i risultati del carico di lavoro sono a rischio, in modo da poter rispondere adeguatamente, se necessario.
- **Attivazione di un avviso quando vengono rilevate delle anomalie nel carico di lavoro**: Attiva un avviso quando vengono rilevate delle anomalie nel carico di lavoro, in modo da poter rispondere adeguatamente, se necessario.
- **Convalida del raggiungimento dei risultati e dell'efficacia dei KPI e dei parametri** : Crea una vista a livello di business delle tue operazioni del carico di lavoro, per stabilire se le esigenze sono soddisfatte e per identificare gli aspetti da migliorare per raggiungere gli obiettivi di business. Convalida l'efficacia dei KPI e dei parametri e rivedili, se necessario.

OPS 9 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:

- **Identificazione degli indicatori chiave delle prestazioni**: Identifica gli indicatori chiave delle prestazioni (KPI) in base all'obiettivo desiderato (ad esempio, fornitura di nuove caratteristiche) e ai risultati dei clienti (ad esempio, casi del servizio clienti). Valuta i KPI per determinare il successo delle operazioni.
- **Definizione dei parametri delle operazioni**: Definisci i parametri delle operazioni per misurare il raggiungimento dei KPI (ad esempio, distribuzioni riuscite e distribuzioni non riuscite). Definisci i parametri delle operazioni per misurare lo stato delle attività operative (ad esempio, tempo medio per rilevare un incidente (MTTD) e tempo medio per il ripristino (MTTR) in seguito a un incidente). Valuta i parametri per stabilire se le operazioni raggiungono i risultati previsti e per comprendere lo stato delle loro attività.
- **Raccolta e analisi dei parametri delle operazioni**: Esegui regolarmente revisioni proattive dei parametri per identificare le tendenze e stabilire dove sono necessarie risposte adeguate.
- **Definizione delle baseline per i parametri delle operazioni**: Definisci le baseline per i parametri in modo da fornire i valori previsti di base per il confronto e l'identificazione delle attività operative con prestazioni basse e alte.

- **Acquisizione dei modelli di attività previsti per le operazioni**: Definisci modelli di attività operative per identificare comportamenti anomali in modo da rispondere in modo appropriato, se necessario.
- **Attivazione di un avviso quando i risultati delle operazioni sono a rischio**: Attiva un avviso quando i risultati delle operazioni sono a rischio in modo da poter rispondere adeguatamente, se necessario.
- **Attivazione di un avviso quando vengono rilevate delle anomalie nelle operazioni**: Attiva un avviso quando vengono rilevate delle anomalie nelle operazioni in modo da poter rispondere adeguatamente, se necessario.
- **Convalida del raggiungimento dei risultati e dell'efficacia dei KPI e dei parametri** : Crea una vista a livello di business delle attività operative, per stabilire se le esigenze sono soddisfatte e per identificare gli aspetti da migliorare per raggiungere gli obiettivi di business. Convalida l'efficacia dei KPI e dei parametri e rivedili, se necessario.

OPS 10 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:

- **Utilizzo di processi per la gestione di eventi, incidenti e problemi**: Predisponi processi per affrontare gli eventi osservati, gli eventi che richiedono un intervento (incidenti) e gli eventi che richiedono un intervento e si verificano nuovamente oppure non possono essere risolti al momento (problemi). Utilizza questi processi per mitigare l'impatto di questi eventi sull'azienda e sui clienti con risposte tempestive e adeguate.
- **Definizione di un processo per ogni avviso**: Predisponi una risposta specifica (runbook o playbook), con un proprietario espressamente identificato, per ogni evento per cui viene generato un avviso. Questo consente di rispondere agli eventi operativi in modo rapido ed efficace, evitando che gli eventi che richiedono un'azione vengano oscurati da notifiche meno importanti.
- **Prioritizzazione degli eventi operativi in base all'impatto aziendale**: Quando più eventi richiedono un intervento, assicurati che quelli più significativi per il business vengano affrontati per primi. Ad esempio, gli impatti possono includere decesso e infortunio, perdite finanziarie o danni alla reputazione o alla fiducia.
- **Definizione dei percorsi di escalation**: Definisci percorsi di escalation nei tuoi runbook e playbook, compresi gli eventi che attivano l'escalation e le procedure di escalation. In particolare, identifica i proprietari per ogni azione per assicurare risposte rapide ed efficaci agli eventi operativi.
- **Abilitazione delle notifiche push**: Informa direttamente gli utenti (ad esempio tramite e-mail o SMS) quando i servizi che utilizzano sono interessati e quando vengono ripristinate le normali condizioni operative, per consentire loro di adottare le misure appropriate.
- **Comunicazione dello stato tramite pannelli di controllo**: Fornisci pannelli di controllo personalizzati in base ai destinatari, ad esempio i team tecnici interni, la dirigenza e i clienti, per comunicare lo stato operativo corrente del business e fornire i parametri desiderati.

- **Automazione delle risposte agli eventi**: Automatizza le risposte agli eventi per ridurre gli errori causati dai processi manuali e assicurare risposte rapide e coerenti.

Evoluzione

OPS 11 In che modo fai evolvere le operazioni?

Dedica tempo e risorse al miglioramento incrementale continuo, per far evolvere l'efficacia e l'efficienza delle tue operazioni.

Best practice:

- **Definizione di un processo per il miglioramento continuo**: Valuta regolarmente e dai priorità alle opportunità di miglioramento per concentrare gli sforzi laddove è possibile ottenere i maggiori vantaggi.
- **Esecuzione di 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 di contributo e le azioni correttive secondo necessità, con modalità adatte al pubblico a cui ti rivolgi.
- **Implementazione di cicli di feedback**: Includi cicli di feedback nelle procedure e nei carichi di lavoro per riuscire a identificare i problemi e gli aspetti da migliorare.
- **Gestione delle conoscenze**: I membri del tuo team dispongono di meccanismi per trovare le informazioni che cercano in modo tempestivo, per accedervi e per verificare che siano attuali e complete. Sono disponibili meccanismi per identificare i contenuti necessari, i contenuti che necessitano di aggiornamento e i contenuti che devono essere archiviati per non essere più richiamati.
- **Definizione dei fattori che promuovono il miglioramento**: Identifica i fattori che promuovono il miglioramento, in modo da valutare e dare priorità alle opportunità.
- **Convalida delle informazioni**: 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.
- **Revisione dei parametri delle operazioni**: Esegui regolarmente un'analisi retrospettiva dei parametri operativi con i partecipanti di vari team da diverse aree del business. Utilizza queste revisioni per identificare opportunità di miglioramento e potenziali linee d'azione e per condividere le conoscenze acquisite.
- **Documentazione e condivisione delle conoscenze acquisite**: Documenta e condividi le conoscenze acquisite durante l'esecuzione delle attività operative, per metterle a frutto internamente e nei vari team.
- **Allocazione di tempo per i miglioramenti**: Dedica tempo e risorse all'interno dei processi per rendere possibile il miglioramento incrementale continuo.

Sicurezza

Sicurezza

SEC 1 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 di AWS e del settore nonché con l'intelligence sulle minacce aiuta a sviluppare il modello di rischio e gli obiettivi di controllo. L'automazione dei processi di sicurezza, i test e la convalida consentono di ricalibrare le operazioni di sicurezza.

Best practice:

- **Carichi di lavoro separati tramite account**: Organizza i carichi di lavoro in account e account di gruppo separati in base alla funzione o a un set di controlli comune invece di ricalcare la struttura organizzativa dell'azienda. Inizia tenendo conto della sicurezza e dell'infrastruttura per consentire alla tua organizzazione di impostare guardrail comuni al crescere dei carichi di lavoro.
- **Protezione dell'account AWS**: Proteggi l'accesso agli account, ad esempio abilitando l'autenticazione MFA, limitando l'uso dell'utente root e configurando i contatti dell'account.
- **Identificazione e convalida degli obiettivi di controllo**: In base ai requisiti di conformità e ai rischi identificati dal modello di rischio, deriva 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.
- **Aggiornamento costante sulle minacce alla sicurezza**: Riconosci i vettori di attacco rimanendo aggiornato sulle minacce alla sicurezza più recenti per definire e implementare controlli appropriati.
- **Aggiornamento costante sulle raccomandazioni di sicurezza**: Tieniti aggiornato sulle raccomandazioni di sicurezza di AWS e del settore, così da revisionare l'assetto di sicurezza del tuo carico di lavoro.
- **Automatizzazione dei test e della convalida dei controlli di sicurezza nelle pipeline**: Stabilisci previsioni e modelli sicuri per i meccanismi di sicurezza testati e convalidati come parte della compilazione, delle pipeline e dei processi. Utilizza strumenti e l'automazione per testare e convalidare tutti i controlli di sicurezza in modo continuo. Ad esempio, scansiona elementi quali immagini di macchine e modelli di infrastrutture come codice per individuare vulnerabilità di sicurezza, irregolarità e deviazioni da una previsione stabilita in ogni fase.
- **Identificazione e assegnazione di priorità ai rischi utilizzando un modello di minaccia**: Utilizza un modello di rischio per identificare e mantenere un registro aggiornato delle potenziali minacce. Classifica le minacce in ordine di priorità e adatta i controlli di sicurezza in modo da prevenirle, rilevarle e affrontarle. Rivedi e mantieni questo approccio nel contesto dell'evoluzione del panorama della sicurezza.

- **Valutazione e implementazione periodiche di nuovi servizi e funzionalità di sicurezza**: AWS e i partner APN rilasciano costantemente nuovi servizi e funzionalità che consentono di aggiornare l'assetto di sicurezza del carico di lavoro.

Gestione di identità e accessi (Identity and Access Management)

SEC 2 Come si gestisce l'autenticazione per persone e macchine?

Ci sono due tipi di identità da gestire quando ci si avvicina all'utilizzo di carichi di lavoro AWS sicuri. Comprendere il tipo di identità necessaria per gestire e concedere l'accesso ti aiuta a garantire che le identità corrette abbiano accesso alle risorse giuste nelle condizioni adeguate. Identità umane: amministratori, sviluppatori, operatori e utenti finali necessitano di un'identità per accedere agli ambienti e alle applicazioni AWS. Si tratta di membri dell'organizzazione o utenti esterni con cui collabori e che interagiscono con le tue risorse AWS tramite browser Web, applicazioni client o strumenti a riga di comando interattivi. Identità di macchine: le applicazioni di servizio, gli strumenti operativi e i carichi di lavoro necessitano di un'identità per effettuare richieste ai servizi AWS, ad esempio per leggere i dati. Queste identità includono macchine in esecuzione nell'ambiente AWS, ad esempio istanze Amazon EC2 o funzioni AWS Lambda. Puoi gestire le identità di macchine anche per soggetti esterni che necessitano dell'accesso. Inoltre, possono esistere macchine al di fuori di AWS che hanno bisogno di accedere al tuo ambiente AWS.

Best practice:

- **Utilizza meccanismi di accesso efficaci**: Imposta la lunghezza minima della password e spiega agli utenti la necessità di evitare password comuni o utilizzate in precedenza. Applica la Multi-Factor Authentication (MFA) con meccanismi software o hardware per garantire un livello aggiuntivo.
- **Utilizza credenziali temporanee**: Richiedi alle identità di acquisire dinamicamente credenziali temporanee. Per le identità della forza lavoro, utilizza AWS Single Sign-On o la federazione con ruoli IAM per accedere agli account AWS. Per le identità di macchine, è necessario utilizzare ruoli IAM anziché chiavi di accesso a lungo termine.
- **Archivia e utilizza i segreti in modo sicuro**: Se le identità della forza lavoro e delle macchine necessitano di segreti (come password per applicazioni di terze parti), memorizzali con rotazione automatica in un servizio specializzato che utilizzi i più recenti standard di settore.
- **Fai affidamento su un provider di identità centralizzato**: Per le identità della forza lavoro, fai affidamento su un provider di identità che ti consente di gestire le identità in un luogo centralizzato. In questo modo puoi creare, gestire e revocare l'accesso da una singola posizione, semplificando la gestione degli accessi. Ciò riduce la necessità di molteplici credenziali e offre l'opportunità di integrarsi con i processi delle risorse umane.
- **Verifica e ruota periodicamente le credenziali**: Quando non puoi fare affidamento sulle credenziali temporanee e devi richiedere credenziali a lungo termine, verificale per assicurarti che siano applicati i controlli prestabiliti (ad esempio, la MFA), siano soggette regolarmente a rotazione e dispongano di un livello di accesso appropriato.

- **Sfrutta i gruppi di utenti e gli attributi**: Inserisci gli utenti con requisiti di sicurezza comuni in gruppi definiti dal provider di identità e metti in atto meccanismi per garantire che gli attributi utente che potrebbero essere utilizzati per il controllo degli accessi (ad esempio, reparto o posizione) siano corretti e aggiornati. Utilizza questi gruppi e attributi, anziché i singoli utenti, per controllare l'accesso. In questo modo è possibile gestire l'accesso centralmente, modificando una volta sola l'appartenenza o gli attributi di un gruppo utente, anziché aggiornare numerose policy individuali quando le esigenze di accesso di un utente cambiano.

SEC 3 Come si gestisce l'autenticazione per persone e macchine?

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

Best practice:

- **Definizione dei requisiti di accesso**: Ogni componente o risorsa del carico di lavoro deve essere accessibile da amministratori, utenti finali o altri componenti. Individua una definizione chiara di chi o cosa deve avere accesso a ciascun componente, quindi scegli il tipo di identità e il metodo di autenticazione e autorizzazione appropriati.
- **Concedi l'accesso con privilegi minimi**: Concedi alle identità soltanto il livello di accesso di cui hanno bisogno, specificando le operazioni che possono effettuare, le risorse AWS su cui possono operare e a quali condizioni. Affidati ai gruppi e agli attributi di identità per impostare dinamicamente 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, quando uno sviluppatore lascia il gruppo, perderà l'accesso a tutte le risorse gestite tramite il gruppo e non sarà necessario modificare le policy di accesso.
- **Determina un processo per l'accesso di emergenza**: Un processo che consente l'accesso di emergenza al carico di lavoro nell'improbabile caso di un problema a un processo automatizzato o a una pipeline. Questo consente di utilizzare criteri di accesso con privilegi minimi, ma garantisce che gli utenti possano ottenere il corretto livello di accesso quando ne hanno bisogno. Ad esempio, puoi stabilire un processo di verifica e approvazione delle richieste degli utenti da parte degli amministratori.
- **Riduci le autorizzazioni in modo continuo**: Man mano che i team e i carichi di lavoro determinano l'accesso di cui hanno bisogno, rimuovi le autorizzazioni che non utilizzano più e stabilisci processi di revisione per applicare le autorizzazioni con privilegi minimi. Monitora e riduci continuamente le identità e le autorizzazioni non utilizzate.
- **Definisci i guardrail per le autorizzazioni della tua organizzazione**: Stabilisci controlli comuni che limitano l'accesso a tutte le identità nella tua organizzazione. Ad esempio, puoi limitare l'accesso a regioni AWS specifiche o impedire agli operatori di eliminare risorse comuni, come ad esempio un ruolo IAM utilizzato per il team di sicurezza centrale.
- **Gestione degli accessi in base al ciclo di vita**: Integra i controlli degli accessi con il ciclo di vita degli operatori e delle applicazioni e con il tuo provider di federazione centralizzata. Ad esempio, rimuovi l'accesso di un utente quando lascia l'organizzazione o cambia ruolo.

- **Analizza l'accesso pubblico e tra account**: Monitora continuamente i risultati che evidenziano l'accesso pubblico e tra account diversi. Limita l'accesso pubblico e l'accesso tra account alle risorse che ne hanno bisogno.
- **Condivi le risorse in modo sicuro**: Regola il consumo di risorse condivise tra account diversi o all'interno della tua AWS Organization. Monitora le risorse condivise e rivedi l'accesso alle stesse.

Rilevamento

SEC 4 In che modo individui ed esamini gli eventi di sicurezza?

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

Best practice:

- **Configurazione della registrazione di log dei servizi e delle applicazioni**: Configura la registrazione di log per tutto il carico di lavoro, inclusi log di applicazioni, di risorse e di servizi AWS. Assicurati ad esempio che AWS CloudTrail, Amazon CloudWatch Logs, Amazon GuardDuty e AWS Security Hub siano abilitati per tutti gli account all'interno dell'organizzazione.
- **Analisi di log, risultati e parametri a livello centrale**: È opportuno raccogliere tutti i log, i parametri e la telemetria a livello centrale e analizzarli automaticamente per rilevare anomalie e indici di attività non autorizzate. Un pannello di controllo consente di accedere facilmente a informazioni sullo stato aggiornate in tempo reale. Ad esempio, assicurati che i log di Amazon GuardDuty e Security Hub vengano inviati a una posizione centrale per avvisi e analisi.
- **Automazione delle risposte agli eventi**: L'utilizzo dell'automazione per analizzare e correggere gli eventi riduce l'impegno e il rischio di errori umani e consente di dimensionare le capacità di analisi. Le revisioni periodiche ti aiuteranno a ottimizzare gli strumenti di automazione e a effettuare un'iterazione costante. Ad esempio, rendi automatiche le risposte agli eventi di Amazon GuardDuty automatizzando la prima fase di verifica, quindi esegui l'iterazione per ridurre gradualmente l'impegno umano.
- **Implementazione di eventi di sicurezza fruibili**: Crea e invia al tuo team avvisi fruibili. Assicurati che includano informazioni pertinenti affinché il team possa intervenire. Ad esempio, verifica che gli avvisi di Amazon GuardDuty e AWS Security Hub siano inviati al team perché vi risponda direttamente oppure che siano inviati agli strumenti di automazione della risposta, con il team tenuto al corrente attraverso messaggi provenienti dal framework di automazione.

Protezione dell'infrastruttura

SEC 5 Come proteggere 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:

- **Creazione di livelli di rete**: Raggruppa i componenti che condividono requisiti di raggiungibilità in vari livelli. Ad esempio, un cluster di database in un VPC senza necessità di accesso a Internet deve essere posizionato in sottoreti senza routing da o verso Internet. In un carico di lavoro serverless che opera senza un VPC, livelli e segmentazione simili con microservizi possono raggiungere lo stesso obiettivo.
- **Controllo del traffico a tutti i livelli**: Applica controlli con un approccio di difesa approfondito sia per il traffico in entrata sia per quello in uscita. Ad esempio, nel caso di Amazon Virtual Private Cloud (VPC), sono previsti gruppi di sicurezza, ACL di rete e sottoreti. Per quanto riguarda AWS Lambda, considera la possibilità di esecuzione nel VPC privato con controlli basati su VPC.
- **Automatizzazione della protezione di rete**: Automatizza i meccanismi di protezione per creare una rete in grado di difendersi da sola grazie alle informazioni sulle minacce e al rilevamento delle anomalie. Ad esempio, introducendo strumenti di rilevamento e prevenzione delle intrusioni in grado di adattarsi alle minacce attuali e di ridurne l'impatto.
- **Implementazione di funzioni di ispezione e protezione**: Ispeziona e filtra il traffico a ogni livello. Ad esempio, utilizza un firewall per applicazioni Web per proteggere da accessi involontari a livello di rete dell'applicazione. Per le funzioni Lambda, strumenti di terze parti possono aggiungere un firewall a livello di applicazione all'ambiente di runtime.

SEC 6 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 ed interne. Le risorse di calcolo includono istanze EC2, container, funzioni di AWS Lambda, servizi di database, dispositivi IoT e altro.

Best practice:

- **Gestione delle vulnerabilità**: Scansiona e correggi frequentemente le vulnerabilità del codice, delle dipendenze e dell'infrastruttura per proteggere da nuove minacce.
- **Riduzione della superficie d'attacco**: Riduci la superficie di attacco attraverso la protezione avanzata dei sistemi operativi e riducendo al minimo i componenti, le librerie e i servizi di consumo esterni in uso.
- **Implementazione di servizi gestiti**: Implementa servizi che gestiscono le risorse, ad esempio Amazon RDS, AWS Lambda e Amazon ECS, per ridurre le attività di manutenzione della sicurezza nell'ambito del modello di responsabilità condivisa.
- **Automatizzazione della protezione delle risorse di calcolo**: Automatizza i meccanismi di protezione delle risorse di calcolo, tra cui la gestione delle vulnerabilità, la riduzione della superficie di attacco e la gestione delle risorse.
- **Concessione del permesso di eseguire azioni a distanza**: Eliminare la possibilità di accesso interattivo riduce il rischio di errore umano e la potenziale configurazione o gestione manuale. Ad esempio, usa un flusso di lavoro per distribuire le istanze EC2 utilizzando l'infrastruttura come codice, dopodiché gestiscile mediante strumenti, invece di consentire l'accesso diretto o attraverso un host bastione.

- **Convalida dell'integrità del software**: Implementa meccanismi (ad esempio la firma del codice) per verificare che il software, il codice e le librerie utilizzati nel carico di lavoro provengano da origini attendibili e non siano stati manomessi.

Protezione dei dati

SEC 7 In che modo classificare 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:

- **Identificazione dei dati all'interno del carico di lavoro**: Le informazioni dovrebbero includere il tipo e la classificazione dei dati, i processi aziendali associati, il proprietario dei dati, i requisiti legali e di conformità applicabili, dove sono archiviati e i controlli risultanti da applicare. Ciò può includere classificazioni per indicare se i dati sono destinati a essere disponibili al pubblico, se i dati sono solo di uso interno, ad esempio informazioni che consentono l'identificazione personale del cliente (PII, Personally Identifiable Information), oppure se i dati riguardano un accesso più limitato, ad esempio relativi alla proprietà intellettuale, dati confidenziali o sensibili e altro ancora.
- **Definizione dei controlli di protezione dei dati**: Proteggi i dati in base al livello di classificazione. Ad esempio, puoi mettere in sicurezza le informazioni classificate come pubbliche utilizzando raccomandazioni pertinenti e allo stesso tempo proteggere i dati sensibili con controlli aggiuntivi.
- **Automatizzazione dell'identificazione e della classificazione**: Automatizza l'identificazione e la classificazione dei dati per ridurre il rischio di errore umano dovuto a interazioni manuali.
- **Definizione della gestione del ciclo di vita dei dati**: La strategia del ciclo di vita definita deve basarsi sul livello di sensibilità e sui requisiti legali e aziendali. Gli aspetti da considerare includono la durata di conservazione dei dati, la distruzione dei dati, la gestione dell'accesso ai dati, la trasformazione dei dati e la condivisione dei dati.

SEC 8 In che modo proteggere i dati inattivi?

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

Best practice:

- **Implementazione della gestione sicura delle chiavi**: Le chiavi di crittografia devono essere conservate in modo sicuro e sottoposte a un rigido controllo degli accessi, ad esempio utilizzando un servizio di gestione come AWS KMS. Valuta l'utilizzo di chiavi diverse e il controllo degli accessi alle chiavi, in combinazione con le policy AWS IAM e delle risorse, per l'adeguamento ai livelli di classificazione dei dati e ai requisiti di separazione.
- **Applicazione della crittografia dei dati inattivi**: Applica i requisiti di crittografia basati sugli standard e sulle raccomandazioni più recenti per favorire la protezione dei dati inattivi.

- **Automatizzazione della protezione dei dati inattivi**: Utilizza strumenti automatizzati per convalidare e applicare la protezione dei dati inattivi in modo continuo; ad esempio verifica che siano presenti solo risorse di storage crittografate.
- **Applicazione del controllo degli accessi**: Applica il controllo degli accessi con privilegi minimi e meccanismi come backup, isolamento e versioning, per favorire la protezione dei dati inattivi. Impedisci agli operatori di concedere l'accesso pubblico ai tuoi dati.
- **Utilizzo di meccanismi per tenere le persone a distanza dai dati**: Evita a tutti gli utenti di accedere direttamente a dati e sistemi sensibili in circostanze operative normali. Fornisci ad esempio un pannello di controllo anziché l'accesso diretto a un datastore per eseguire query. Se non vengono utilizzate le pipeline CI/CD, determina quali controlli e processi sono necessari per fornire in modo adeguato un meccanismo di accesso di tipo break-glass normalmente disabilitato.

SEC 9 Come proteggere i dati in transito?

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

Best practice:

- **Implementazione della gestione sicura delle chiavi e dei certificati**: Memorizza chiavi di crittografia e certificati in modo sicuro e ruotali a intervalli di tempo adeguati, applicando un controllo degli accessi severo; puoi utilizzare ad esempio un servizio di gestione dei certificati come AWS Certificate Manager.
- **Applicazione della crittografia dei dati in transito**: Applica i tuoi requisiti di crittografia definiti in base ad appropriati standard e raccomandazioni in modo da soddisfare i requisiti aziendali, legali e di conformità.
- **Automatizzazione del rilevamento degli accessi indesiderati ai dati**: Utilizza strumenti come Amazon GuardDuty per rilevare automaticamente i tentativi di spostamento dei dati al di fuori di limiti definiti in base al livello di classificazione dei dati, ad esempio per rilevare un trojan che sta copiando i dati in una rete sconosciuta o non attendibile utilizzando il protocollo DNS.
- **Autenticazione delle comunicazioni di rete**: Verifica l'identità delle comunicazioni utilizzando protocolli che supportano l'autenticazione, ad esempio Transport Layer Security (TLS) o IPsec.

Risposta agli incidenti

SEC 10 In che modo è possibile prevedere gli eventi, rispondervi e risolverli?

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'organizzazione.

Best practice:

- **Identificazione del personale chiave e delle risorse esterne**: Identifica personale, risorse e requisiti legali interni ed esterni che potrebbero aiutare l'organizzazione a rispondere a un incidente.
- **Sviluppo di piani di gestione degli incidenti**: Crea piani che ti aiutino a rispondere a un incidente, comunicare durante lo stesso e ripristinare in seguito le risorse. Ad esempio, puoi avviare un piano di risposta agli incidenti con gli scenari più probabili per il carico di lavoro e l'organizzazione. Includi il modo in cui gestiresti la comunicazione e l'escalation internamente ed esternamente.
- **Preparazione di funzionalità forensi**: Identifica e prepara capacità di indagine forensi idonee, tra cui specialisti esterni, strumenti e automazione.
- **Automatizzazione della capacità di contenimento**: Automatizza il contenimento di un incidente e il successivo ripristino per ridurre i tempi di risposta e l'impatto sull'organizzazione.
- **Preassegnazione dell'accesso**: Assicurati che il team di risposta agli incidenti disponga del corretto accesso preassegnato ad AWS per ridurre i tempi di verifica fino al ripristino.
- **Distribuzione anticipata degli strumenti**: Assicurati che il personale addetto alla sicurezza disponga degli strumenti giusti pre-distribuiti in AWS per ridurre i tempi di verifica fino al ripristino.
- **Esecuzione di giornate di gioco**: Esercitati regolarmente con giornate di gioco (simulazioni) di risposta agli incidenti, integra i concetti appresi nei piani di gestione degli incidenti e cerca di migliorare costantemente.

Affidabilità

Fondamenti

REL 1 Come si gestiscono quote e vincoli di servizio?

Per le architetture di carichi di lavoro basate sul cloud, esistono quote di servizio (definite anche come restrizioni dei servizi). Queste quote sono presenti per evitare di effettuare accidentalmente il provisioning di più risorse di quelle necessarie e limitare i tassi di richiesta sulle operazioni API in modo da proteggere i servizi da un uso illecito. Esistono anche vincoli di risorse, ad esempio la velocità con cui è possibile trasferire i bit su un cavo in fibra ottica o la quantità di storage su un disco fisico.

Best practice:

- **Consapevolezza su quote e vincoli di servizio**: Conosci le quote predefinite e le richieste di aumento delle quote per l'architettura del carico di lavoro. Inoltre, sai quali vincoli delle risorse, ad esempio disco o rete, sono potenzialmente influenti.
- **Gestione delle quote di servizio in più account e regioni**: Se utilizzi più account AWS o regioni AWS, assicurati di richiedere le quote appropriate in tutti gli ambienti in cui vengono eseguiti i carichi di lavoro di produzione.

- **Rispetto di quote e vincoli di servizio fissi mediante l'architettura**: Considera le quote di servizio immutabili e le risorse fisiche e progetta per evitare che queste compromettano l'affidabilità.
- **Monitoraggio e gestione delle quote**: Valuta il tuo utilizzo potenziale e aumenta le quote in modo appropriato per una crescita pianificata dell'utilizzo.
- **Automazione della gestione delle quote**: Implementa strumenti per ricevere avvisi quando le soglie stanno per essere raggiunte. Utilizzando le API di AWS Service Quotas, puoi automatizzare le richieste di aumento delle quote.
- **Assicurarsi che vi sia un intervallo sufficiente tra le quote attuali e l'utilizzo massimo per consentire eventuali failover**: Quando una risorsa presenta un errore, può continuare a essere conteggiata ai fini del raggiungimento delle quote fino a quando non viene terminata correttamente. Assicurati che le quote coprano la sovrapposizione di tutte le risorse non riuscite con sostituzioni prima che le risorse non riuscite vengano terminate. Nel calcolo di questo intervallo dovresti considerare un errore nella zona di disponibilità.

REL 2 Come si pianifica la topologia di rete?

I carichi di lavoro sono spesso presenti in più ambienti. Questi includono più ambienti cloud (sia pubblicamente accessibili 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:

- **Utilizzo di una connettività di rete a disponibilità elevata per gli endpoint pubblici del carico di lavoro**: Questi endpoint e il routing verso di essi devono essere altamente disponibili. Per ottenere questo risultato, utilizza DNS ad alta disponibilità, reti di distribuzione di contenuti (CDN), API Gateway, bilanciamento del carico o proxy inversi.
- **Effettua il provisioning di connettività ridondante tra reti private nel cloud e negli ambienti in locale**: Utilizza più connessioni AWS Direct Connect (DX) o tunnel VPN tra reti private distribuite separatamente. Utilizza più ubicazioni DX per un'elevata disponibilità. Se utilizzi più regioni AWS, garantisci la ridondanza in almeno due di esse. È possibile valutare le appliance AWS Marketplace che terminano le VPN. Se utilizzi appliance di AWS Marketplace, distribuisci le istanze ridondanti per la disponibilità elevata in diverse zone di disponibilità.
- **Verificare che l'allocazione delle sottoreti IP consenta l'espansione e la disponibilità**: Gli intervalli di indirizzi IP di Amazon VPC devono essere abbastanza grandi da soddisfare i requisiti del carico di lavoro, tra cui la fattorizzazione nella futura espansione e l'allocazione di indirizzi IP alle sottoreti nelle zone di disponibilità. Sono inclusi sistemi di bilanciamento del carico, istanze EC2 e applicazioni basate su container.
- **Preferire topologie hub-and-spoke rispetto a mesh molti-a-molti**: Se più di due spazi di indirizzi di rete (ad esempio, VPC e reti in locale) sono connessi tramite peering VPC, AWS Direct Connect o VPN, utilizza un modello hub-and-spoke, come quello fornito da AWS Transit Gateway.

- **Applicazione di intervalli di indirizzi IP privati non sovrapposti in tutti gli spazi con indirizzi privati a cui sono connessi**: Gli intervalli di indirizzi IP di ogni VPC non devono sovrapporsi quando collegati in peering o connessi tramite VPN. Analogamente, è necessario evitare conflitti di indirizzi IP tra un VPC e ambienti in locale o con altri provider di servizi cloud utilizzati. Bisogna inoltre disporre di un modo per allocare gli intervalli di indirizzi IP privati quando necessario.

Architettura del carico di lavoro

REL 3 Come si progetta l'architettura del servizio di carico di lavoro?

Creazione di 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.

Best practice:

- **Scelta del tipo di segmentazione del carico di lavoro**: L'architettura monolitica deve essere evitata. Al contrario, è consigliabile scegliere tra SOA e microservizi. Quando effettui ogni scelta, bilancia i vantaggi con le complessità: ciò che è giusto per un nuovo prodotto che corre al primo lancio è diverso da ciò che necessita un carico di lavoro creato per ricalibrare le risorse dall'inizio. I vantaggi dell'utilizzo di segmenti più piccoli includono maggiore agilità, flessibilità organizzativa e scalabilità. Le complessità includono un eventuale aumento della latenza, un debug più complesso e un maggiore carico operativo
- **Creazione di servizi focalizzati su domini e funzionalità aziendali specifici**: SOA crea servizi con funzioni ben delineate definite dalle esigenze aziendali. I microservizi utilizzano modelli di dominio e contesto delimitato per restringere ulteriormente questa operazione, in modo che ogni servizio esegua una sola operazione. Focalizzarsi su funzionalità specifiche consente di differenziare i requisiti di affidabilità dei diversi servizi e mirare agli investimenti in modo più specifico. Un problema aziendale conciso e l'associazione di un piccolo team a ciascun servizio facilitano il dimensionamento dell'organizzazione.
- **Fornitura di contratti di servizio per API**: I contratti di servizio sono accordi documentati tra i team sull'integrazione dei servizi e includono una definizione API leggibile dal computer, limiti di velocità e aspettative di prestazioni. Una strategia di controllo delle versioni consente ai client di continuare a utilizzare l'API esistente e migrare le applicazioni all'API più recente quando sono pronte. La distribuzione può avvenire in qualsiasi momento, purché il contratto non venga violato. Il team del fornitore di servizi può utilizzare lo stack tecnologico scelto per soddisfare il contratto API. Analogamente, l'utente del servizio può utilizzare la propria tecnologia.

REL 4 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 in 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 impediscono gli errori e migliorano il tempo medio tra errori (MTBF).

Best practice:

- **Identificazione del tipo di sistema distribuito necessario**: I sistemi distribuiti hard real-time richiedono risposte che devono essere fornite in modo sincrono e rapido, mentre i sistemi soft real-time hanno una finestra temporale più generosa di minuti o più per la risposta. I sistemi offline gestiscono le risposte tramite elaborazione in batch o asincrona. I sistemi distribuiti hard real-time hanno i requisiti di affidabilità più severi.
- **Implementazione di dipendenze "loosely coupled"**: Le dipendenze come sistemi di accodamento, sistemi di streaming, flussi di lavoro e sistemi di bilanciamento del carico sono "loosely coupled" (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.
- **Rendere 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. Per eseguire questa operazione, i client possono inviare richieste API con un token di idempotenza: viene utilizzato lo stesso token ogni volta che si ripete la richiesta. Un'API del servizio idempotente utilizza il token per restituire una risposta identica a quella restituita la prima volta che la richiesta è stata completata.
- **Esecuzione di un lavoro costante**: I sistemi possono fallire quando si verificano modifiche rapide e di grandi dimensioni nel carico. Ad esempio, un sistema di controllo dello stato che monitora lo stato 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 dello stato esegue un lavoro costante con modifiche rapide e di piccole dimensioni.

REL 5 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 consentono ai carichi di lavoro di affrontare stress o guasti, recuperare più rapidamente e mitigare l'impatto di tali problemi. Il risultato è un miglioramento del tempo medio di ripristino (MTTR).

Best practice:

- **Implementazione del degrado elegante per trasformare le dipendenze forti applicabili in dipendenze deboli**: Quando le dipendenze di un componente non sono integre, il componente stesso può comunque funzionare, anche se in modo degradato. Ad esempio, quando una chiamata di dipendenza non riesce, utilizza invece una risposta statica predeterminata.
- **Richieste di throttling**: Si tratta di un modello di mitigazione per rispondere a un aumento imprevisto della domanda. Alcune richieste vengono rispettate, ma quelle che superano

un limite definito vengono rifiutate e restituiscono un messaggio che indica che sono state sottoposte a throttling. L'aspettativa per i client è che si ritirino e abbandonino la richiesta o riprovino a una velocità più lenta.

- **Controllo e limitazione delle chiamate di ripetizione**: Utilizza il backoff esponenziale per eseguire nuovi tentativi dopo intervalli progressivamente più lunghi. Introduci il jitter per randomizzare gli intervalli di ripetizione e limitare il numero massimo di tentativi.
- **Errore rapido e limitazione delle code**: Se il carico di lavoro non è in grado di rispondere correttamente a una richiesta, restituisce rapidamente un errore. Ciò consente il rilascio delle risorse associate a una richiesta e permette al servizio di recuperare se le risorse sono in esaurimento. Se il carico di lavoro è in grado di rispondere correttamente, ma la frequenza delle richieste è troppo elevata, utilizza una coda per eseguire il buffer delle richieste. Tuttavia, non consentire code lunghe che possono comportare l'elaborazione di richieste obsolete a cui il client ha già rinunciato.
- **Impostazione dei timeout dei client**: Imposta i timeout in modo appropriato, verificali sistematicamente e non fare affidamento sui valori predefiniti poiché sono generalmente troppo alti
- **Rendere i servizi stateless laddove possibile**: I servizi dovrebbero non richiedere lo stato oppure 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. In questo modo i server possono essere sostituiti a piacimento senza compromettere la disponibilità. Amazon ElastiCache o Amazon DynamoDB sono ottime destinazioni per lo stato di offload.
- **Implementazione di leve di emergenza**: Si tratta di processi rapidi che possono mitigare l'impatto della disponibilità sul carico di lavoro. Possono essere utilizzati in assenza di una causa principale. Una leva di emergenza ideale riduce a zero il carico cognitivo dei resolver fornendo criteri di attivazione e disattivazione completamente deterministici. Le leve di esempio includono il blocco di tutto il traffico del robot o la distribuzione di una risposta statica. Le leve sono spesso manuali, ma possono anche essere automatizzate.

Gestione delle modifiche

REL 6 Come monitorare le risorse del carico di lavoro?

I log e i parametri sono strumenti molto efficaci per ottenere informazioni sullo stato del tuo carico di lavoro. È possibile configurare il carico di lavoro in modo da monitorare i log e i parametri e inviare notifiche quando vengono superate le soglie o si verificano eventi significativi. Il monitoraggio consente al carico di lavoro di riconoscere quando vengono superate le soglie di prestazioni basse o si verificano errori, in modo che possa essere ripristinato automaticamente di rimando.

Best practice:

- **Monitoraggio di tutti i componenti per il carico di lavoro (generazione)**: Monitora i componenti del carico di lavoro con Amazon CloudWatch o con strumenti di terze parti. Monitoraggio dei servizi AWS con Personal Health Dashboard

- **Definizione e calcolo dei parametri (aggregazione)**: Archivia i dati di log e applica filtri, se necessario, per calcolare i parametri, ad esempio i conteggi di un evento di log specifico o la latenza calcolata dai timestamp degli eventi di log
- **Invia notifiche (elaborazione e avvisi in tempo reale)**: Le organizzazioni che devono essere messe al corrente ricevono le notifiche nel caso si verifichino eventi significativi
- **Automatizza le risposte (elaborazione e avvisi in tempo reale)**: Utilizza l'automazione per agire quando viene rilevato un evento; ad esempio, per sostituire i componenti guasti
- **Archiviazione e analisi**: Raccogli i file di log e le cronologie dei parametri e analizzali per ottenere informazioni più ampie sulle tendenze e sui carichi di lavoro
- **Esecuzione di revisioni periodiche**: Controlla frequentemente l'implementazione del monitoraggio del carico di lavoro e aggiornalo in base a eventi e modifiche significativi
- **Monitoraggio del tracciamento end-to-end delle richieste attraverso il sistema**: Utilizza AWS X-Ray o strumenti di terze parti per consentire agli sviluppatori di analizzare ed eseguire il debug di sistemi distribuiti in modo più semplice per comprendere l'andamento delle prestazioni delle loro applicazioni e dei relativi servizi sottostanti

REL 7 Come si progetta il carico di lavoro per adattarsi ai cambiamenti della domanda?

Un carico di lavoro scalabile fornisce elasticità per aggiungere o rimuovere risorse automaticamente, in modo che vi sia una stretta corrispondenza con la domanda attuale in un dato momento.

Best practice:

- **Utilizzo dell'automazione per l'acquisizione o il dimensionamento delle risorse**: Quando sostituisci risorse danneggiate o ridimensioni il carico di lavoro, automatizza il processo utilizzando servizi AWS gestiti, come Amazon S3 e AWS Auto Scaling. Puoi anche utilizzare strumenti di terze parti ed SDK AWS per automatizzare il dimensionamento.
- **Ottieni le risorse quando viene rilevata la compromissione di un carico di lavoro**: All'occorrenza, ridimensiona le risorse in modo reattivo se la disponibilità è influenzata per ripristinare la disponibilità del carico di lavoro.
- **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.
- **Esecuzione di un test di carico sul carico di lavoro**: Adotta un metodo di test del carico per misurare se l'attività di dimensionamento soddisfa i requisiti del carico di lavoro.

REL 8 In che modo implementare le modifiche?

Per distribuire nuove funzionalità e garantire 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:

- **Uso di runbook per attività standard come la distribuzione**: I runbook sono le fasi predefinite utilizzate per ottenere risultati specifici. Utilizza i runbook per eseguire attività standard, o manualmente o automaticamente. Alcuni esempi includono la distribuzione di un carico di lavoro, l'applicazione di patch o la realizzazione di modifiche DNS.
- **Esegui test funzionali come parte integrante della distribuzione**: I test funzionali vengono eseguiti come parte integrante della distribuzione automatizzata. Se non vengono soddisfatti i criteri di esito positivo, la pipeline viene arrestata o ripresa dall'inizio.
- **Esegui test di resilienza come parte integrante della distribuzione**: I test di resilienza (che rientrano nell'ingegneria del caos) vengono eseguiti nell'ambito della pipeline di distribuzione automatizzata in un ambiente di pre-produzione.
- **Distribuisci utilizzando un'infrastruttura immutabile**: Si tratta di 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.
- **Distribuisci le modifiche tramite automazione**: Le distribuzioni e l'applicazione di patch sono automatizzate per eliminare l'impatto negativo.

Gestione degli errori

REL 9 In che modo eseguire il backup dei dati?

Esegui il backup dei dati, delle applicazioni e della configurazione per soddisfare i tuoi requisiti relativi agli obiettivi di tempo di ripristino (recovery time objective, RTO) e agli obiettivi di punto di ripristino (recovery point objective, RPO).

Best practice:

- **Identificazione di tutti i dati di cui è necessario eseguire il backup oppure riprodurre dei dati dalle origini** : Amazon S3 può essere utilizzato come destinazione di backup per più origini dati. I servizi AWS come Amazon EBS, Amazon RDS e Amazon DynamoDB hanno funzionalità integrate per la creazione di backup. È anche possibile utilizzare software di backup di terze parti. In alternativa, se per soddisfare gli RPO è possibile riprodurre i dati da altre origini, il backup potrebbe non essere necessario.
- **Protezione e codifica dei backup**: Rileva l'accesso tramite autenticazione e autorizzazione, ad esempio AWS IAM, e individua un'eventuale compromissione dell'integrità dei dati utilizzando la crittografia.
- **Esecuzione del backup dei dati in automatico**: Configura i backup in modo che vengano eseguiti automaticamente in base a una pianificazione periodica o mediante modifiche nel set di dati. Le istanze RDS, i volumi EBS, le tabelle DynamoDB e gli oggetti S3 possono essere configurati per il backup automatico. È anche possibile utilizzare soluzioni AWS Marketplace o soluzioni di terze parti.
- **Esegui periodicamente il ripristino dei dati per verificare l'integrità e i processi di backup**: Esegui un test di ripristino per verificare che l'implementazione del processo di backup soddisfi gli obiettivi di tempo di ripristino (recovery time objective, RTO) e gli obiettivi di punto di ripristino (recovery point objective, RPO).

REL 10 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:

- **Distribuzione del carico di lavoro in diversi luoghi**: 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.
- **Ripristino automatico dei componenti vincolati a una singola posizione**: Se i componenti del carico di lavoro possono essere eseguiti solo in una singola zona di disponibilità o in un data center locale, è necessario implementare la capacità di eseguire una ricostruzione completa del carico di lavoro entro gli obiettivi di ripristino definiti.
- **Utilizzo di architetture paratie**: Come le paratie su una nave, questo modello garantisce che un guasto sia contenuto a un piccolo sottoinsieme di richieste/utenti, in modo che il numero di richieste danneggiate sia limitato e la maggior parte possa continuare senza errori. Le paratie per i dati sono in genere chiamate partizioni o shard, mentre le paratie per i servizi sono note come celle.

REL 11 Come si progetta il carico di lavoro affinché resista ai guasti dei componenti?

I carichi di lavoro con requisiti di disponibilità elevata e MTTR (Mean Time To Recovery) basso devono essere progettati per garantire la resilienza.

Best practice:

- **Monitoraggio di tutti i componenti del carico di lavoro per rilevare i guasti**: Monitora continuamente lo stato del carico di lavoro, in modo che tu e i tuoi sistemi automatizzati siate consapevoli del deterioramento o del guasto completo non appena questi si verificano. Monitora gli indicatori chiave di prestazioni (KPI) in base al valore aziendale.
- **Failover su risorse integre in posizioni non danneggiate**: Assicurati che se si verifica un guasto in una posizione, i dati e le risorse provenienti da posizioni integre possono continuare a servire le richieste. Ciò è più semplice per i carichi di lavoro multi-zona, poiché i servizi AWS, come Elastic Load Balancing e AWS Auto Scaling, aiutano a distribuire il carico tra le zone di disponibilità. Per i carichi di lavoro multi-regione, questa operazione è più complicata. Ad esempio, le repliche di lettura tra regioni consentono di distribuire i dati in più regioni AWS, ma è comunque necessario promuovere la replica di lettura per dominare e indirizzare il traffico verso di essa in caso di guasto di una posizione primaria. Amazon Route 53 e AWS Global Accelerator possono anche aiutare a instradare il traffico tra regioni AWS.
- **Automatizzazione del risanamento a tutti i livelli**: Al rilevamento di un guasto, utilizza funzionalità automatizzate per eseguire azioni da correggere.
- **Utilizzo della stabilità statica per evitare un comportamento bimodale**: Si ha un comportamento bimodale 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 ha esito negativo. Devi invece creare carichi di lavoro che siano staticamente stabili e operino in una sola modalità. In questo caso, effettua il provisioning di istanze sufficienti in ciascuna zona di disponibilità per gestire il carico di lavoro se una zona di disponibilità è stata rimossa, quindi utilizza Elastic Load Balancing o i controlli dello stato di Amazon Route 53 per spostare il carico dalle istanze danneggiate.

- **Invio di notifiche quando gli eventi influiscono sulla disponibilità**: Le notifiche vengono inviate al rilevamento di eventi significativi, anche se il problema causato dall'evento è stato risolto automaticamente.

REL 12 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 garantire il funzionamento corretto e offrire la resilienza prevista.

Best practice:

- **Utilizzo dei playbook per analizzare gli errori**: Abilita 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 qualsiasi fase del processo vengono utilizzati per stabilire i passaggi da intraprendere successivamente fino all'identificazione o alla risoluzione del problema.
- **Esecuzione di 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 imprevisti. 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.
- **Test dei requisiti funzionali**: Includono test delle unità e test di integrazione che convalidano la funzionalità richiesta.
- **Test dei requisiti di dimensionamento e prestazioni**: Includono il test del carico per verificare che il carico di lavoro soddisfi i requisiti di dimensionamento e prestazioni.
- **Test della resilienza tramite l'utilizzo dell'ingegneria del caos**: Esegui test che inseriscono regolarmente guasti negli ambienti di pre-produzione e produzione. Ipotizza il modo in cui il carico di lavoro reagirà al guasto, quindi confronta la tua ipotesi con i risultati del test ed esegui l'iterazione se non corrispondono. Assicurati che il test di produzione non influisca sugli utenti.
- **Esecuzione regolare di giornate di gioco**: Utilizza le giornate di gioco per provare regolarmente le procedure di errore il 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 i test di produzione non influiscano sugli utenti.

REL 13 Come si pianifica il disaster recovery?

Avere backup e componenti del carico di lavoro ridondanti in loco è l'inizio della strategia di disaster recovery. RTO e RPO sono i tuoi obiettivi per il ripristino della disponibilità. 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.

Best practice:

- **Definizione degli obiettivi di ripristino in caso di downtime e perdita di dati**: Il carico di lavoro ha un Recovery Time Objective (RTO) e Recovery Point Objective (RPO).
- **Utilizzo di strategie di ripristino definite per conseguire gli obiettivi di ripristino**: Per conseguire gli obiettivi è stata definita una strategia di Disaster Recovery (DR).
- **Esecuzione di test sull'implementazione del disaster recovery per convalidare l'implementazione**: Esegui regolarmente il test di failover su DR per assicurarti che siano soddisfatti RTO e RPO.
- **Gestione della deviazione di configurazione nel sito o nella regione del DR**: Assicurati che l'infrastruttura, i dati e la configurazione soddisfino le esigenze del sito o nella regione del DR. Ad esempio, controlla che le AMI e le quote di servizio siano aggiornate.
- **Automatizzazione del ripristino**: Utilizza AWS o strumenti di terze parti per automatizzare il ripristino del sistema e instradare il traffico verso il sito o la regione DR.

Efficienza delle prestazioni

Selezione

PERF 1 In che modo selezioni l'architettura più performante?

Spesso sono necessari molteplici approcci per ottenere prestazioni ottimali in un carico di lavoro. I sistemi Well-Architected utilizzano soluzioni multiple e funzionalità diverse per migliorare le prestazioni.

Best practice:

- **Identificazione dei servizi e delle risorse disponibili**: Scopri tutte le informazioni sull'ampia gamma di servizi e risorse disponibili nel cloud. Identifica quali servizi e opzioni di configurazione sono pertinenti per il tuo carico di lavoro e studia come utilizzarli per raggiungere prestazioni ottimali.
- **Definizione di un processo per le scelte architetturali**: Affidati all'esperienza e alle competenze del cloud o utilizza risorse esterne, come casi di utilizzo pubblicati, documentazione pertinente o whitepaper, per definire un processo per scegliere risorse e servizi. È necessario definire un processo che incoraggi la sperimentazione e il benchmarking con i servizi che potrebbero essere utilizzati nel tuo carico di lavoro.
- **Requisiti di costo dei fattori nelle decisioni** : I carichi di lavoro spesso hanno requisiti di costo per il funzionamento. Utilizza i controlli dei costi interni per selezionare le dimensioni e i tipi di risorse in base alle necessità previste in termini di risorse.

- **Uso di policy o architetture di riferimento**: Massimizza le prestazioni e l'efficienza valutando le policy interne e le architetture di riferimento esistenti e sfrutta la tua analisi per selezionare servizi e configurazioni per il carico di lavoro.
- **Utilizzo delle linee guida del fornitore di servizi cloud o di un partner appropriato**: Utilizza le risorse del fornitore di servizi cloud, come solutions architect, servizi professionali o un partner appropriato per orientare le tue decisioni. Queste risorse possono aiutarti a rivedere e migliorare l'architettura per ottenere prestazioni ottimali.
- **Benchmarking dei carichi di lavoro esistenti**: Esegui il benchmarking delle prestazioni di un carico di lavoro esistente per comprendere le sue prestazioni sul cloud. Utilizza i dati raccolti da questi benchmark per orientare le decisioni architetturali.
- **Esecuzione di un test di carico sul carico di lavoro**: Distribuisci l'architettura del carico di lavoro più recente nel cloud utilizzando tipologie e dimensioni di risorse diverse. Monitora la distribuzione per acquisire parametri delle prestazioni che identificano colli di bottiglia o capacità in eccesso. Utilizza queste informazioni sulle prestazioni per progettare o migliorare la tua architettura e la selezione delle risorse.

PERF 2 In che modo selezioni la soluzione di calcolo?

La soluzione di calcolo ottimale 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 consentire funzioni diverse per migliorare le prestazioni. Selezionare la soluzione di calcolo sbagliata per un'architettura può portare a una riduzione dell'efficienza delle prestazioni.

Best practice:

- **Valutazione delle opzioni di elaborazione disponibili**: Studia e comprendi le caratteristiche di prestazione delle opzioni relative all'elaborazione disponibili. Comprendi il modo in cui funzionano le istanze, i container e le funzioni e quali siano i vantaggi e gli svantaggi che comportano per il tuo carico di lavoro.
- **Identificazione delle opzioni di configurazione dell'elaborazione disponibili**: Comprendi in che modo le varie opzioni completano il tuo carico di lavoro e quali opzioni di configurazione sono le migliori per il tuo sistema. Esempi di tali opzioni includono la famiglia di istanze, le dimensioni, le caratteristiche (GPU, I/O), le dimensioni delle funzioni, le istanze di container, multitenancy o singola, e così via.
- **Raccolta dei parametri relativi all'elaborazione**: Uno dei migliori modi per comprendere le prestazioni dei tuoi sistemi di calcolo è registrare e tracciare l'utilizzo effettivo di varie risorse. Questi dati possono essere utilizzati per determinare in modo più accurato i requisiti delle risorse.
- **Definizione della configurazione richiesta in base al corretto dimensionamento**: Analizza le varie caratteristiche di prestazione del tuo carico di lavoro e come queste sono correlate a memoria, rete e utilizzo della CPU. Utilizza questi dati per scegliere le risorse che meglio corrispondono al profilo del tuo carico di lavoro. Ad esempio, un carico di lavoro a memoria elevata, come un database, potrebbe essere servito meglio dalla famiglia di istanze r. Al contrario, un carico di lavoro con picchi di prestazioni può trarre maggiori vantaggi da un sistema di container elastici.

- **Utilizzo dell'elasticità disponibile delle risorse**: 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. Se combinato con parametri relativi all'elaborazione, un carico di lavoro può rispondere automaticamente a questi cambiamenti e utilizzare la gamma di risorse più opportuna per raggiungere il suo obiettivo.
- **Rivalutazione delle esigenze di elaborazione sulla base dei parametri**: Utilizza i parametri a livello di sistema per identificare il comportamento e i requisiti del tuo carico di lavoro nel tempo. Valuta le esigenze del tuo carico di lavoro confrontando le risorse disponibili con tali requisiti e apporta modifiche al tuo ambiente di elaborazione per soddisfare al meglio il profilo del carico di lavoro. Ad esempio, nel corso del tempo si potrebbe osservare che un sistema utilizza molta più memoria di quanto si pensasse inizialmente, e trasferirlo a una famiglia o una dimensione di istanze diversa potrebbe migliorarne sia le prestazioni sia l'efficienza.

PERF 3 In che modo selezioni la soluzione di storage?

La soluzione di storage ottimale per un sistema varia in base a fattori quali: tipo di metodo di accesso (blocco, file od oggetto), schemi di accesso (casuali o sequenziali), throughput necessario, frequenza di accesso (online, offline, archivio), frequenza di aggiornamento (WORM, dinamico) e vincoli di disponibilità e durata. I sistemi Well-Architected utilizzano più soluzioni di storage e consentono funzionalità diverse per migliorare le prestazioni e utilizzare le risorse in modo efficiente.

Best practice:

- **Identificazione delle caratteristiche e i requisiti di storage**: Studia le diverse caratteristiche (ad esempio possibilità di condivisione, dimensioni dei file, dimensioni della cache, schemi di accesso, latenza, throughput e persistenza dei dati) necessarie per selezionare i servizi più adatti al carico di lavoro, ad esempio storage a oggetti, storage a blocchi, storage a file o storage dell'istanza.
- **Valutazione delle opzioni di configurazione disponibili**: Valuta le varie caratteristiche e opzioni di configurazione e il modo in cui sono correlate allo storage. Comprendi dove e come utilizzare Provisioned IOPS, SSD, storage magnetico, storage a oggetti, storage di archiviazione o storage temporaneo per ottimizzare lo spazio di storage e le prestazioni del tuo carico di lavoro.
- **Decisioni basate su schemi e parametri di accesso**: Scegli i sistemi di storage in base agli schemi di accesso del carico di lavoro e configurali determinando il modo in cui il carico di lavoro accede ai dati. Aumenta l'efficienza dello storage scegliendo lo storage di oggetti anziché lo storage a blocchi. Configura le opzioni di storage in funzione dei tuoi schemi di accesso ai dati.

PERF 4 In che modo selezioni la soluzione di database?

La soluzione di database ottimale per un determinato sistema può variare in base ai requisiti di disponibilità, coerenza, tolleranza della partizione, latenza, durata, scalabilità e capacità di query. Molti sistemi utilizzano diverse soluzioni di database per vari sottosistemi e consentono funzionalità differenti per migliorare le prestazioni. La selezione della soluzione e delle funzionalità errate del database per un sistema può ridurre l'efficienza delle prestazioni.

Best practice:

- **Comprendi le caratteristiche dei dati**: Comprendi le diverse caratteristiche dei dati nel tuo carico di lavoro. Determina se il carico di lavoro necessita di transazioni, in che modo interagisce con i dati e quali sono le sue esigenze in termini di prestazioni. Utilizza tali dati per selezionare l'approccio di database con le prestazioni migliori per il tuo carico di lavoro (ad esempio storage con database relazionali, chiave-valore NoSQL, documento, colonnare, grafi, serie temporali o in memoria).
- **Valutazione delle opzioni disponibili**: Valuta i servizi e le opzioni di storage disponibili come parte del processo di selezione per i meccanismi di storage del tuo carico di lavoro. Comprendi come e quando utilizzare un determinato servizio o sistema per lo storage dei dati. Scopri le opzioni di configurazione disponibili in grado di ottimizzare le prestazioni o l'efficienza del database, ad esempio Provisioned IOPS, risorse di memoria ed elaborazione e memorizzazione nella cache.
- **Raccolta e registrazione dei parametri delle prestazioni del database**: Utilizza strumenti, librerie e sistemi che registrano misure delle prestazioni relative alle prestazioni del database. Per esempio, misura il numero di transazioni per secondo, query lente o la latenza del sistema introdotta al momento dell'accesso al database. Utilizza questi dati per comprendere le prestazioni dei sistemi di database.
- **Scelta dello storage dei dati in base agli schemi di accesso**: Utilizza gli schemi di accesso del carico di lavoro per decidere quali servizi e tecnologie utilizzare. Per esempio, utilizza un database relazionale per i carichi di lavoro che necessitano di transazioni, o uno store chiave-valore che fornisce un throughput maggiore ma anche una lettura finale consistente, ove applicabile.
- **Ottimizzazione dello storage dei dati in base agli schemi e ai parametri di accesso**: Utilizza caratteristiche delle prestazioni e schemi di accesso che ottimizzano il modo in cui i dati vengono archiviati o interrogati al fine di ottenere le migliori prestazioni possibili. Misura il modo in cui le ottimizzazioni come l'indicizzazione, la distribuzione delle chiavi, la progettazione dei data warehouse o le strategie di memorizzazione nella cache influenzano le prestazioni del sistema o la sua efficienza nel complesso.

PERF 5 In che modo configuri la tua soluzione di rete?

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 in locale, determinano le opzioni di posizione. Questi vincoli possono essere compensati con le edge location o la collocazione delle risorse.

Best practice:

- **Comprensione dell'impatto della rete sulla performance**: Analizza e comprendi come le decisioni relative alla rete influenzano le prestazioni di rete. Ad esempio, la latenza della rete spesso e volentieri influisce sull'esperienza dell'utente, e l'utilizzo di protocolli errati può minare la capacità di rete con un sovraccarico eccessivo.
- **Valutazione delle 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.

- **Scelta di una connettività dedicata o una VPN di dimensioni adeguate ai carichi di lavoro ibridi**: Quando è richiesta comunicazione in locale, assicurati di disporre di una larghezza di banda adeguata alle prestazioni del carico di lavoro. In base ai requisiti di larghezza di banda, una singola connessione dedicata o una singola VPN potrebbe non essere sufficiente, rendendo pertanto necessaria l'abilitazione del bilanciamento del carico del traffico su più connessioni.
- **Sfruttamento del bilanciamento del carico e dell'offloading della crittografia**: 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, gestire e instradare il traffico in modo efficiente.
- **Scelta dei 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.
- **Scelta della posizione del carico di lavoro in base ai requisiti di rete**: Utilizza le opzioni di posizione nel cloud disponibili per ridurre la latenza di rete o migliorare il throughput. Utilizza regioni AWS, zone di disponibilità, gruppi di collocazione e edge location come Outposts, regioni locali e Wavelength per ridurre la latenza di rete o migliorare il throughput.
- **Ottimizzazione della configurazione di rete in base ai parametri**: Usa i dati raccolti e analizzati per prendere decisioni informate riguardo l'ottimizzazione della configurazione della tua rete. Misura l'impatto di tali cambiamenti e usa le misurazioni per prendere decisioni future.

Revisione

PERF 6 In che modo fai evolvere il carico di lavoro per sfruttare le nuove versioni?

Quando si progettano carichi di lavoro, le opzioni tra cui scegliere sono limitate. Tuttavia, nel tempo diventano disponibili nuove tecnologie e nuovi approcci che potrebbero migliorare le prestazioni.

Best practice:

- **Mantenersi aggiornati sui nuovi servizi e sulle nuove risorse**: Valuta i modi per migliorare le prestazioni man mano che nuovi servizi, modelli di progettazione e offerte di prodotti diventano disponibili. Studia come le novità potrebbero migliorare le prestazioni o aumentare l'efficienza del carico di lavoro tramite una valutazione ad hoc, una discussione interna o un'analisi esterna.
- **Definizione di 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.

- **Evoluzione delle prestazioni del carico di lavoro nel corso del tempo**: Come organizzazione, utilizza le informazioni raccolte durante il processo di valutazione per gestire attivamente l'adozione di nuovi servizi o risorse quando diventano disponibili.

Monitoraggio

PERF 7 In che modo monitori le tue risorse per assicurarti che abbiano le giuste prestazioni?

Le prestazioni del sistema possono peggiorare nel tempo. Monitora le prestazioni del sistema per identificare l'eventuale riduzione delle prestazioni e rimediare a fattori interni o esterni, come il sistema operativo o il carico dell'applicazione.

Best practice:

- **Registrazione dei parametri relativi alle prestazioni**: Utilizza un servizio di monitoraggio e osservazione per registrare i parametri correlati alle prestazioni. Ad esempio, registra le transazioni di database, le query lente, la latenza I/O, il throughput delle richieste HTTP, la latenza del servizio o altri dati chiave.
- **Analisi dei parametri in caso di eventi o incidenti**: In risposta a nel corso di un evento o un incidente, utilizza pannelli di controllo o report di monitoraggio per comprendere e diagnosticare l'impatto. Queste viste forniscono informazioni sulle parti del carico di lavoro le cui prestazioni non raggiungono i livelli previsti.
- **Individuazione degli indicatori chiave di prestazione (KPI) per misurare le prestazioni del carico di lavoro**: Identifica i KPI che mostrano se il carico di lavoro sta funzionando come previsto. Un carico di lavoro basato su API, ad esempio, può utilizzare la latenza di risposta complessiva come indicazione delle prestazioni complessive, mentre per un sito di e-commerce un KPI valido può essere il numero di acquisti andati a buon fine.
- **Utilizzo del monitoraggio per generare notifiche basate su allarmi**: Avvalendoti degli indicatori chiave di prestazione (KPI) relativi alle prestazioni che hai identificato, utilizza un sistema di monitoraggio che genera automaticamente allarmi quando queste misurazioni sono al di fuori dei limiti previsti.
- **Analisi dei parametri a intervalli regolari**: Come manutenzione ordinaria o in risposta a eventi o incidenti, esamina quali parametri vengono raccolti. Stabilisci quali di questi parametri sono fondamentali per risolvere i problemi e quali altri parametri aggiuntivi, se monitorati, potrebbero contribuire a identificare, affrontare o prevenire i problemi.
- **Monitoraggio e allarmi proattivi**: Utilizza indicatori chiave di prestazioni (KPI), in combinazione con sistemi di monitoraggio e allarmi, per risolvere in modo proattivo i problemi correlati alle prestazioni. 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 coloro che possono intervenire. Ad esempio, puoi implementare un sistema in grado di prevedere i valori attesi per i KPI e di inviare allarmi qualora essi oltrepassino determinate soglie, oppure uno strumento che arresta o esegue automaticamente il rollback delle distribuzioni nel caso in cui i valori dei KPI si discostino dai valori attesi.

Compromessi

PERF 8 Come si utilizzano i compromessi per migliorare le prestazioni?

Quando si progettano soluzioni, determinare i compromessi ti consente di selezionare un approccio ottimale. Spesso è possibile migliorare le prestazioni accettando compromessi in termini di coerenza, durata e spazio a favore di tempo e latenza.

Best practice:

- **Definizione delle aree in cui le prestazioni sono più importanti**: 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.
- **Studio dei servizi e dei modelli di progettazione**: Ricerca e analizza i vari servizi e modelli di progettazione che permettono di migliorare le prestazioni del carico di lavoro. Nell'ambito dell'analisi, identifica gli elementi sui quali potresti accettare compromessi per ottenere prestazioni più elevate. Ad esempio, l'utilizzo di un servizio di cache può contribuire a ridurre il carico sui sistemi di database, tuttavia richiede una certa quantità di progettazione per l'implementazione di cache sicure o l'eventuale introduzione di consistenza finale in alcune aree.
- **Identificazione dell'impatto dei compromessi sui clienti e sull'efficienza**: Quando valuti i miglioramenti correlati alle prestazioni, determina quali scelte avranno un 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 della sua natura finale influirà sui clienti.
- **Misurazione dell'impatto dei miglioramenti delle prestazioni**: Quando vengono apportate modifiche per migliorare le prestazioni, valuta i parametri e i dati raccolti. Utilizza queste informazioni per determinare l'impatto che il miglioramento delle prestazioni ha avuto sul carico di lavoro, sui suoi componenti e sui clienti. Queste misurazioni permettono di capire quali sono i miglioramenti ottenuti dai compromessi applicati e aiutano a stabilire se si sono verificati eventuali effetti collaterali negativi.
- **Scelta di più strategie relative alle prestazioni**: Se possibile, utilizza più strategie per migliorare le prestazioni. Scegli, ad esempio, strategie come la memorizzazione dei dati nella cache per evitare eccessive chiamate di rete o dei database, l'utilizzo di repliche di lettura per i motori di database al fine di migliorare i tassi di lettura, lo sharding o la compressione dei dati, ove possibile, per ridurre i volumi, e il buffering e lo streaming dei risultati man mano che diventano disponibili per evitare blocchi.

Ottimizzazione dei costi

Esercizio della gestione finanziaria del cloud

COST 1 Come implementi la gestione finanziaria nel cloud?

L'implementazione della gestione finanziaria del cloud consente alle organizzazioni di conseguire un valore aggiunto e il successo finanziario ottimizzando i costi e l'utilizzo e ricalibrando le risorse in AWS.

Best practice:

- **Stabilire una funzione di ottimizzazione dei costi**: Crea un team responsabile di stabilire e mantenere la consapevolezza dei costi in tutta l'organizzazione. Il team richiede collaboratori dai ruoli finanziari, tecnologici e aziendali in tutta l'organizzazione.
- **Stabilire una partnership tra team finanziari e tecnologici**: Coinvolgi i team finanziari e tecnologici nelle discussioni su costi e utilizzo in tutte le fasi del tuo approccio al 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.
- **Stabilire budget e previsioni per il cloud**: Adatta i processi di previsione e di budget 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 di business o una combinazione di entrambi.
- **Implementazione della consapevolezza dei costi nei processi dell'organizzazione**: Implementa la consapevolezza dei costi in processi nuovi o esistenti che influiscono sull'utilizzo e sfrutta i processi esistenti per la consapevolezza dei costi. Implementa la consapevolezza dei costi nella formazione dei dipendenti.
- **Invio di report e notifiche sull'ottimizzazione dei costi**: Configura Budget AWS per fornire notifiche su costi e utilizzo rispetto ai target. Organizza riunioni regolari per analizzare l'efficienza dei costi di questo carico di lavoro e promuovere una cultura attenta ai costi.
- **Monitoraggio proattivo dei costi**: Implementa strumenti e pannelli di controllo per monitorare i costi in modo proattivo per il carico di lavoro. Non guardare solo i costi e le categorie quando ricevi le notifiche. Questo aiuta a identificare gli andamenti positivi e a promuoverli attraverso la tua organizzazione.
- **Mantieniti aggiornato sulle nuove versioni dei servizi**: Consultati regolarmente con gli esperti o con i partner APN per valutare quali servizi e caratteristiche offrono un costo inferiore. Rivedi i blog AWS e altre fonti di informazione.

Consapevolezza delle spese e dell'utilizzo

COST 2 In che modo gestisci l'utilizzo?

Stabilisci policy e meccanismi per assicurarti di sostenere costi adeguati mentre raggiungi gli obiettivi. Utilizzando un approccio di controllo e bilanciamento reciproco, è possibile innovare senza spendere troppo.

Best practice:

- **Sviluppo di politiche basate sui requisiti della tua organizzazione**: Sviluppa politiche che definiscono come le risorse vengono gestite dalla tua organizzazione. Le policy devono coprire gli aspetti dei costi relativi alle risorse e ai carichi di lavoro, compresa la creazione, la modifica e la disattivazione nel ciclo di vita delle risorse.

- **Implementazione di obiettivi e target**: Implementa obiettivi di costi e utilizzo per il carico di lavoro. Gli obiettivi forniscono indicazioni alla tua organizzazione su costi e utilizzo e i target forniscono risultati misurabili per i tuoi carichi di lavoro.
- **Implementazione di una struttura di account**: Implementa una struttura di account che si adatta alla tua organizzazione. Questo aiuta a ripartire e gestire i costi in tutta la tua organizzazione.
- **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 si applica ai servizi AWS e a soluzioni di terze parti.
- **Implementazione dei controlli di costo**: Implementa controlli basati sulle policy dell'organizzazione e gruppi e ruoli definiti. Questi garantiscono che i costi siano sostenuti solo in base ai requisiti dell'organizzazione, ad esempio, controllano l'accesso alle regioni o ai tipi di risorse con le policy IAM.
- **Registrazione 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.

COST 3 In che modo monitori l'utilizzo e il costo?

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

Best practice:

- **Configurazione di fonti di informazione dettagliate**: Configura i report costi e utilizzo AWS e la granularità oraria di Cost Explorer per fornire informazioni dettagliate su costi e utilizzo. Configura il carico di lavoro per far sì che le voci di log vengano registrate per ogni risultato aziendale distribuito.
- **Identificazione delle categorie di attribuzione dei costi**: Identifica le categorie dell'organizzazione che possono essere utilizzate per allocare i costi all'interno della tua organizzazione.
- **Definizione dei parametri dell'organizzazione**: Definisci i parametri dell'organizzazione necessari per questo carico di lavoro. I parametri esemplificativi di un carico di lavoro sono i report dei clienti prodotti o le pagine web disponibili ai clienti.
- **Configurazione degli strumenti di fatturazione e di gestione dei costi**: Configura Cost Explorer AWS e Budget AWS in linea con le policy della tua organizzazione.
- **Aggiunta di informazioni sull'organizzazione a costi e utilizzo**: Definisci uno schema di applicazione di tag basato sull'organizzazione, attributi del carico di lavoro e categorie di allocazione dei costi. Implementa l'applicazione di tag su tutte le risorse. Utilizza Cost Categories per raggruppare i costi e l'utilizzo in base agli attributi dell'organizzazione.
- **Allocazione dei costi in base ai parametri del carico di lavoro**: Alloca i costi del carico di lavoro in base ai parametri o ai risultati aziendali per misurare l'efficienza dei costi del carico di lavoro. Implementa un processo per analizzare il Report costi e utilizzo AWS con Amazon Athena, che può fornire informazioni e capacità di chargeback.

COST 4 In che modo disattivi le risorse?

Implementa il controllo del cambiamento e la gestione delle risorse dall'inizio del progetto alla fine del ciclo di vita. In questo modo, puoi chiudere o interrompere le risorse non utilizzate per ridurre gli sprechi.

Best practice:

- **Monitoraggio delle risorse lungo il loro ciclo di vita**: Definisci e implementa un metodo per monitorare le risorse e le loro associazioni con i sistemi lungo il loro ciclo di vita. Puoi usare l'applicazione di tag per identificare il carico di lavoro o la funzione della risorsa.
- **Implementazione di un processo di disattivazione**: Implementa un processo per identificare e disattivare le risorse orfane.
- **Disattivazione delle risorse**: Disattivazione delle risorse attivate da eventi come audit periodici o modifiche relative all'utilizzo. La disattivazione viene in genere eseguita periodicamente ed è manuale o automatizzata.
- **Disattivazione delle risorse in modo automatico**: Progetta il tuo carico di lavoro in modo da gestire con eleganza l'interruzione delle risorse, identificando e disattivando le risorse non critiche, le risorse non necessarie o quelle a basso utilizzo.

Risorse convenienti

COST 5 In che modo valuti i costi quando selezioni i servizi?

Amazon EC2, Amazon EBS e Amazon S3 sono servizi AWS del blocco predefinito. I servizi gestiti, come Amazon RDS e Amazon DynamoDB, sono servizi AWS di livello superiore o di livello applicazione. 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:

- **Identificazione dei requisiti dell'organizzazione per i 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.
- **Analisi di tutti i componenti di questo carico di lavoro**: Assicurati 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.
- **Esecuzione di un'analisi accurata di ciascun componente**: Considera il costo complessivo per l'organizzazione di ogni componente. Considera il costo di proprietà totale tenendo conto dei costi operativi e di gestione, soprattutto quando si utilizzano i servizi gestiti. L'attività di revisione deve riflettere i potenziali benefici; ad esempio, il tempo speso per l'analisi è proporzionale al costo dei componenti.
- **Selezione di 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.

Nei casi in cui il software con licenza sia obbligatorio, evita le licenze legate ad attributi arbitrari, ad esempio CPU, e cerca le licenze legate all'output o ai risultati. Il costo di queste licenze si ridimensiona in base ai vantaggi che offrono.

- **Selezione dei componenti del 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. Questo include l'utilizzo di servizi a livello applicativo e gestiti come Amazon RDS, Amazon DynamoDB, Amazon SNS e Amazon SES per ridurre il costo complessivo dell'organizzazione. Utilizza serverless e container per l'elaborazione, come AWS Lambda, Amazon S3 per i siti web statici e Amazon ECS. 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.
- **Esecuzione di un'analisi dei costi per diversi valori di utilizzo 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 all'utilizzo previsto, garantisci che il carico di lavoro rimanga conveniente per tutta la sua durata.

COST 6 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:

- **Esecuzione della modellizzazione dei costi**: Identifica i requisiti dell'organizzazione ed esegui la modellizzazione dei costi del carico di lavoro e di ciascuno dei suoi componenti. Esegui attività di analisi comparativa per il carico di lavoro in base ai diversi carichi previsti e confronta i costi. L'attività di modellazione deve riflettere i potenziali benefici, ad esempio, il tempo speso è proporzionale al costo dei componenti.
- **Selezione del tipo e della dimensione delle risorse in base ai dati**: Seleziona la dimensione o il tipo di risorsa in base ai dati relativi al carico di lavoro e alle caratteristiche delle risorse, ad esempio, elaborazione, memoria, throughput o scrittura intensiva. Questa selezione è tipicamente effettuata utilizzando una versione precedente del carico di lavoro (ad esempio una versione in locale), utilizzando la documentazione o altre fonti di informazione sul carico di lavoro.
- **Selezione automatica del tipo e della dimensione della risorsa in base ai parametri**: Utilizza i parametri del carico di lavoro in esecuzione per selezionare la dimensione e il tipo giusti per ottimizzare i costi. Fornisci adeguatamente throughput, dimensionamento e storage per servizi come Amazon EC2, Amazon DynamoDB, Amazon EBS (PIOPS), Amazon RDS, Amazon EMR e networking. Questa operazione può essere eseguita con un loop di feedback, ad esempio l'auto scaling o tramite codice personalizzato nel carico di lavoro.

COST 7 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:

- **Esecuzione di un'analisi del modello di prezzo**: 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 del carico di lavoro utilizzando la funzione di raccomandazione di AWS Cost Explorer.
- **Implementazione delle regioni in base al costo**: La determinazione dei prezzi delle risorse può essere diversa in ciascuna regione. La considerazione del costo della regione garantisce il pagamento del prezzo complessivo più basso per questo carico di lavoro
- **Selezione di contratti di terze parti con termini convenienti**: Gli accordi e i termini convenienti assicurano che i costi di questi servizi siano ridimensionati in base ai vantaggi che offrono. Seleziona gli accordi e i prezzi che si ridimensionano quando forniscono ulteriori vantaggi alla tua organizzazione.
- **Implementazione di modelli di determinazione dei prezzi per tutti i componenti del 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. L'istanza on demand viene utilizzata 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.
- **Esecuzione dell'analisi del modello di prezzo a livello di account principale**: Utilizza Savings Plans di Cost Explorer e le raccomandazioni sulle istanze riservate per eseguire analisi periodiche a livello di account principale per ottenere sconti a fronte di impegni.

COST 8 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:

- **Esecuzione della modellizzazione 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.
- **Selezione dei componenti per ottimizzare il costo di trasferimento dei dati**: Tutti i componenti sono selezionati e l'architettura è progettata per ridurre i costi di trasferimento dei dati. Questo include l'utilizzo di componenti come l'ottimizzazione WAN e le configurazioni Multi-AZ
- **Implementazione dei servizi per ridurre il costo di trasferimento dei dati**: Implementa servizi per ridurre il trasferimento dei dati, ad esempio, utilizzando un CDN come Amazon CloudFront per fornire contenuti agli utenti finali, livelli di cache utilizzando Amazon ElastiCache o utilizzando AWS Direct Connect invece della VPN per la connettività ad AWS.

Gestione delle risorse di domanda e offerta

COST 9 Come gestisci la domanda e fornisci le risorse?

Per avere un carico di lavoro con costo e prestazioni bilanciate, assicurati che venga utilizzato tutto ciò per cui paghi ed evita le istanze molto sottoutilizzate. Un parametro di utilizzo distorto, in qualsiasi delle suddette direzioni, ha un impatto negativo sull'organizzazione, sia per i costi operativi (basse prestazioni a causa di un utilizzo eccessivo) che per le spese AWS sprecate (a causa di un provisioning eccessivo).

Best practice:

- **Analisi della domanda del carico di lavoro**: Analizza la domanda del carico di lavoro nel tempo. Assicurati 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, il tempo speso è proporzionale al costo del carico di lavoro.
- **Implementazione di un buffer o del throttling per gestire la domanda**: Buffering e throttling modificano la domanda sul carico di lavoro, attenuando eventuali picchi. Implementa il throttling quando i client eseguono nuovi tentativi. Implementa il buffering per archiviare la richiesta e rinviare l'elaborazione a un secondo momento. Assicurati che le esecuzioni di throttling e buffering siano progettate in modo che i client ricevano una risposta nel tempo richiesto.
- **Fornitura dinamica delle risorse**: Le risorse sono fornite in modo pianificato. La pianificazione può essere basata sulla domanda, ad esempio tramite l'auto scaling, oppure sul tempo, quando la domanda è prevedibile e le risorse sono fornite in base al tempo. Questi metodi comportano il minor numero possibile di sovra o sotto-provisioning.

Ottimizzazione nel tempo

COST 10 In che modo valuti i nuovi servizi?

Poiché AWS rilascia nuovi servizi e funzionalità, è consigliabile rivedere le decisioni correnti sull'architettura per garantire che continuino a essere le più convenienti.

Best practice:

- **Sviluppo di 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'attività di revisione deve riflettere i potenziali benefici; ad esempio, i carichi di lavoro principali o i carichi di lavoro con un valore superiore al 10% della fattura sono rivisti trimestralmente, mentre i carichi di lavoro inferiori al 10% sono rivisti annualmente.
- **Valutazione e analisi regolare del carico di lavoro**: I carichi di lavoro esistenti vengono regolarmente rivisti secondo i processi definiti.