

Principio della sostenibilità



Principio della sostenibilità: Framework AWS Well-Architected

Copyright © 2024 Amazon Web Services, Inc. and/or its affiliates. All rights reserved.

Amazon's trademarks and trade dress may not be used in connection with any product or service that is not Amazon's, in any manner that is likely to cause confusion among customers, or in any manner that disparages or discredits Amazon. All other trademarks not owned by Amazon are the property of their respective owners, who may or may not be affiliated with, connected to, or sponsored by Amazon.

Table of Contents

Riassunto e introduzione	i
Introduzione	1
Sostenibilità del cloud	3
Il modello di responsabilità condivisa	4
Sostenibilità del cloud	4
Sostenibilità nel cloud	5
Sostenibilità tramite il cloud	5
Principi di progettazione per l'affidabilità nel cloud	6
Processo di miglioramento	8
Scenario di esempio	9
Identifica gli obiettivi in termini di miglioramento	9
Risorse	9
Valuta miglioramenti specifici	10
Metriche proxy	10
Metriche aziendali	11
Indicatori chiave di prestazione	11
Stima del miglioramento	11
Stima dei miglioramenti	12
Assegna le priorità e pianifica i miglioramenti	13
Testa e convalida i miglioramenti	14
Implementazione delle modifiche in produzione	16
Misurazione dei risultati e replica dei successi	16
Sostenibilità come requisito non funzionale	19
Best practice per la sostenibilità nel cloud	21
Selezione delle regioni	21
SUS01-BP01 Scelta della Regione in base alle esigenze aziendali e agli obiettivi di sostenibilità.	21
Allineamento alla domanda	23
SUS02-BP01 Scala dinamicamente l'infrastruttura dei carichi di lavoro	24
SUS02-BP02 Allineamento degli SLA agli obiettivi di sostenibilità	27
SUS02-BP03 Interruzione della creazione e della manutenzione di risorse inutilizzate	29
SUS02-BP04 Ottimizzazione del posizionamento geografico dei carichi di lavoro in base ai requisiti di rete	31

SUS02-BP05 Ottimizzazione delle risorse dei membri del team in base alle attività eseguite	34
SUS02-BP06 Implementazione del buffering o della limitazione (della larghezza di banda della rete) per ridurre la curva della domanda	36
Software e architettura	39
SUS03-BP01 Ottimizzazione del software e delle architetture per processi asincroni e pianificati	39
SUS03-BP02 Rimozione o rifattorizzazione dei componenti dei carichi di lavoro con un utilizzo ridotto o assente	42
Ottimizzazione delle aree di codice che consumano la maggior parte del tempo o delle risorse	44
SUS03-BP04 Ottimizzazione dell'impatto su dispositivi e apparecchiature	46
SUS03-BP05 Uso dei modelli e le architetture software che meglio supportano l'accesso ai dati e i modelli di archiviazione	49
Gestione dati	51
SUS04-BP01 Implementazione di una policy di classificazione dei dati	52
SUS04-BP02 Utilizzo di tecnologie che supportano l'accesso ai dati e i modelli di archiviazione	53
SUS04-BP03 Utilizzo delle policy per gestire il ciclo di vita dei set di dati	58
SUS04-BP04 Utilizzo dell'elasticità e dell'automazione per espandere lo storage a blocchi o il file system	61
SUS04-BP05 Eliminazione dei dati ridondanti o non necessari	63
SUS04-BP06 Utilizzo di file system condivisi o archiviazione per accedere a dati comuni	65
SUS04-BP07 Riduzione al minimo dello spostamento di dati tra reti	67
SUS04-BP08 Backup dei dati solo quando sono difficili da ricreare	69
Hardware e servizi	71
SUS05-BP01 Utilizzo della quantità minima di hardware per soddisfare le esigenze aziendali	72
SUS05-BP02 Utilizzo di tipi di istanze con il minimo impatto	74
SUS05-BP03 Utilizzo dei servizi gestiti	77
SUS05-BP04 Ottimizzazione dell'uso degli acceleratori di calcolo basati su hardware	79
Processo e cultura	81
SUS06-BP01 Adozione di metodi che consentano di introdurre rapidamente migliorie in tema di sostenibilità	82
SUS06-BP02 Aggiornamento del carico di lavoro	84
SUS06-BP03 Aumento dell'utilizzo degli ambienti di costruzione	86

SUS06-BP04 Utilizzo di device farm gestite per i test	88
Conclusione	90
Collaboratori	91
Approfondimenti	92
Revisioni del documento	93
Avvisi	94
AWS Glossary	95

Pilastro della sostenibilità – Framework AWS Well-Architected

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

Questo whitepaper tratta del pilastro della sostenibilità del Framework Amazon Web Services (AWS) Well-Architected. Offre principi di progettazione, linee guida operative, best practice, compromessi potenziali e piani di miglioramento che puoi usare per soddisfare gli obiettivi di sostenibilità per i carichi di lavoro AWS.

Introduzione

Il Framework AWS Well-Architected aiuta a comprendere i pro e i contro delle decisioni prese durante la creazione dei carichi di lavoro in AWS. Utilizzando il Framework, scoprirai le best practice architetturali per progettare e gestire carichi di lavoro affidabili, sicuri, efficienti, convenienti e sostenibili nel Cloud AWS. Permette di misurare in modo coerente le architetture secondo le best practice e identificare le aree da migliorare. Carichi di lavoro ben progettati aumentano notevolmente la capacità di supportare i risultati aziendali.

Il Framework si basa su sei principi:

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

Questo documento si focalizza sul principio della sostenibilità e, nell'ambito della sostenibilità, si focalizza più precisamente sulla sostenibilità aziendale. Questo documento è rivolto a chi ricopre ruoli nell'ambito della tecnologia, ad esempio ai Chief Technology Officer (CTO), ai progettisti, agli sviluppatori e ai membri dei team operativi.

Grazie a questo documento, comprenderai le attuali raccomandazioni e strategie di AWS da utilizzare durante la progettazione di architetture cloud incentrandole sulla sicurezza. Adottando le strategie

illustrate in questo documento, puoi sviluppare architetture che ottimizzano l'efficienza e riducono gli sprechi.

Sostenibilità del cloud

Per sostenibilità si intende l'impatto economico, ambientale e sociale delle proprie attività aziendali sul lungo termine. La [Commissione mondiale per l'ambiente e lo sviluppo delle Nazioni Unite](#) definisce lo sviluppo sostenibile come "uno sviluppo che soddisfa le esigenze del presente senza compromettere la capacità delle generazioni future di soddisfare le proprie esigenze". La tua organizzazione o azienda potrebbe avere un impatto negativo sull'ambiente con emissioni di anidride carbonica dirette o indirette, rifiuti non riciclabili e danni alle risorse condivise come l'acqua pulita.

Nella fase di sviluppo di carichi di lavoro cloud, adottare un comportamento sostenibile significa essere consapevoli dell'impatto dei servizi utilizzati, quantificare tale impatto per l'intero ciclo di vita del carico di lavoro e applicare principi di progettazione e best practice per ridurre tale impatto. Questo documento si focalizza sull'impatto ambientale, soprattutto in termini di uso ed efficienza delle fonti energetiche, leve importanti che gli architetti usano per definire interventi diretti mirati a ridurre lo sfruttamento delle risorse.

Quando ti focalizzi sugli impatti ambientali, devi capire come questi impatti vengono in genere presi in considerazione e gli impatti successivi sulle emissioni della tua azienda. Il [Greenhouse Gas Protocol](#) organizza le emissioni di anidride carbonica nei seguenti ambiti di applicazione, oltre a fornire esempi pertinenti di emissioni all'interno di ciascun contesto per un provider cloud come AWS:

- Ambito di applicazione 1: Tutte le emissioni dirette provenienti dalle attività di un'organizzazione o sotto il suo controllo. Ad esempio, l'uso di combustibile da parte dei generatori di backup dei data center.
- Ambito di applicazione 2: Emissioni indirette provenienti dall'elettricità acquistata e usata per alimentare data center e altre infrastrutture. Ad esempio, emissioni derivanti dalla generazione di energia a fini commerciali.
- Ambito di applicazione 3: Tutte le emissioni indirette di attività di un'organizzazione da fonti che non controlla. Esempi di AWS includono emissioni relative alla costruzione di data center e alla produzione e al trasporto di hardware IT distribuito in data center.

Dal punto di vista di un cliente AWS, le emissioni dei carichi di lavoro in esecuzione su AWS vengono considerate emissioni indirette e sono parte delle emissioni dell'Ambito di applicazione 3. Ogni carico di lavoro distribuito genera una frazione delle emissioni AWS totali di ognuno degli ambiti di applicazione precedenti. La quantità effettiva varia in base al carico di lavoro e dipende da fattori diversi, tra cui i servizi AWS utilizzati, l'energia consumata da tali servizi, l'intensità di emissione di

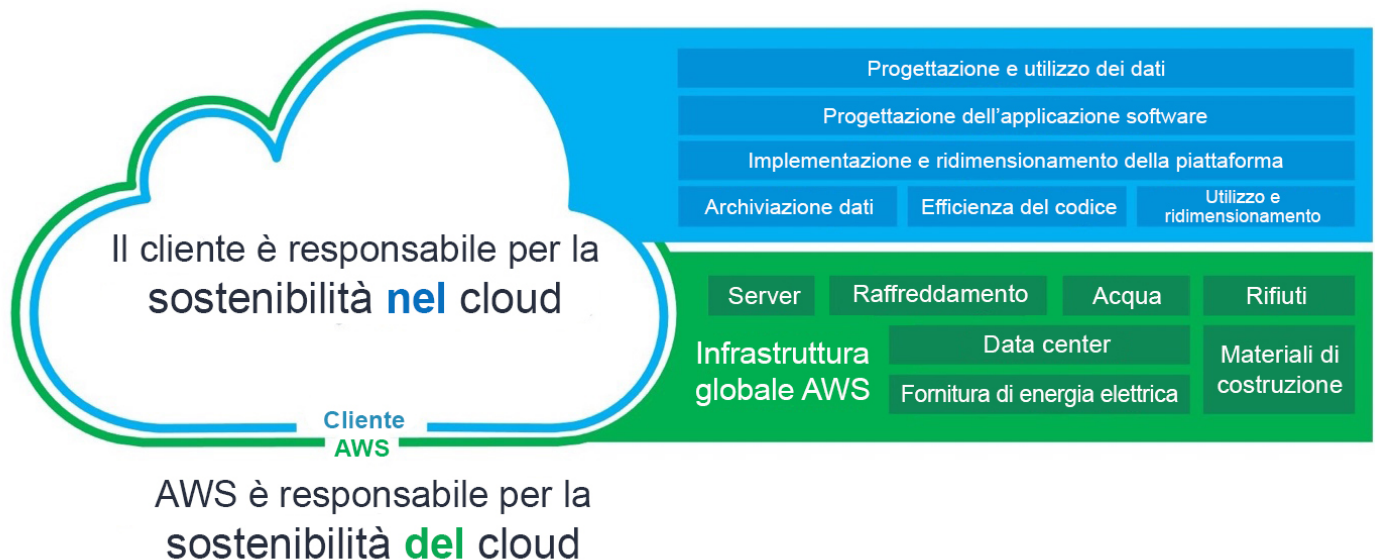
anidride carbonica delle griglie elettriche che alimentano i data center AWS nei punti in cui sono in esecuzione e l'approvvigionamento di AWS di energia rinnovabile.

Questo documento descrive innanzitutto un modello di responsabilità condivisa per la sostenibilità aziendale poi fornisce best practice architettoniche per ridurre al minimo l'impatto dei carichi di lavoro, diminuendo le risorse totali richieste per la loro esecuzione nei data center AWS.

Il modello di responsabilità condivisa

La sostenibilità ambientale è una responsabilità condivisa tra i clienti e AWS.

- AWS è responsabile dell'ottimizzazione della responsabilità del cloud: offrire un'infrastruttura condivisa ed efficiente, amministrare correttamente l'acqua e sfruttare energia rinnovabile.
- I clienti sono responsabili della sostenibilità nel cloud: ottimizzano i carichi di lavoro e l'utilizzo delle risorse, nonché riducono al minimo la quantità di risorse richieste per la distribuzione per i carichi di lavoro.



Modello di responsabilità condivisa

Sostenibilità del cloud

I provider cloud hanno un impatto inferiore in termini di emissioni e offrono una gestione dell'energia più efficiente rispetto alle alternative on-premise tradizionali, poiché investono in energia e in

tecnologie di raffreddamento efficienti, gestiscono gruppi di server validi dal punto di vista dell'energia e ottengono tassi di utilizzo dei server più elevati. I carichi di lavoro nel cloud riducono l'impatto sfruttando le risorse condivise, ad esempio reti, sistemi di alimentazione, raffreddamento e strutture fisiche. Puoi trasferire i tuoi carichi di lavoro nel cloud per avvalerti di tecnologie più efficienti non appena diventano disponibili e utilizzare servizi basati sul cloud per trasformare i tuoi carichi di lavoro per una migliore sostenibilità.

Risorse

- [Opportunità di ridurre le emissioni migrando al Amazon Web Services](#)
- [AWS offre soluzioni di sostenibilità](#)

Sostenibilità nel cloud

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

Sostenibilità tramite il cloud

Oltre a ridurre l'impatto dei carichi di lavoro distribuiti, puoi usare Cloud AWS per eseguire carichi di lavoro progettati per rispondere alle tue sfide di sostenibilità a più ampio raggio. Esempi di queste sfide sono la riduzione delle emissioni, l'abbassamento dei consumi energetici, il riciclo dell'acqua o la diminuzione degli sprechi in altre aree della tua azienda o della tua organizzazione.

La sostenibilità mediante il cloud è quando usi la tecnologia AWS per risolvere una sfida più ampia in termini di sostenibilità. Ad esempio, puoi usare un servizio di machine learning come [Amazon Monitron](#) per rilevare comportamenti anomali nei macchinari industriali. Grazie ai dati rilevati puoi eseguire una manutenzione preventiva per ridurre il rischio di incidenti ambientali causati da errori inaspettati delle apparecchiature e avere la certezza che i macchinari possano continuare a operare al massimo dell'efficienza.

Principi di progettazione per l'affidabilità nel cloud

Applica questi principi di progettazione quando crei i tuoi carichi di lavoro nel cloud per ottimizzare la sostenibilità e ridurre l'impatto.

- **Comprendi il tuo impatto:** Misura l'impatto del tuo carico di lavoro cloud e definisci il suo impatto futuro. Nella tua analisi includi ogni fonte di impatto: quelle derivanti dall'uso dei prodotti da parte dei tuoi clienti e quelle derivanti dalla rimozione e dal ritiro finali dal mercato. Confronta l'output di produzione e l'impatto totale dei tuoi carichi di lavoro cloud, partendo dall'analisi di risorse ed emissioni richieste per unità di lavoro. Usa questi dati per definire indicatori chiave di prestazione (KPI), capire come migliorare la produttività, riducendo al tempo stesso l'impatto, e stimare l'impatto delle modifiche proposte nel tempo.
- **Stabilisci obiettivi di sostenibilità:** Per ogni carico di lavoro cloud stabilisci obiettivi di sostenibilità a lungo termine, come, ad esempio, ridurre le risorse di calcolo e di archiviazione richieste per ciascuna transazione. Modella il ritorno sugli investimenti finalizzati alle migliori in materia di sostenibilità per i carichi di lavoro esistenti e offri ai proprietari le risorse di cui hanno bisogno per investire negli obiettivi di sostenibilità. Pianifica lo sviluppo e progetta i tuoi carichi di lavoro in modo che la crescita comporti un impatto meno intenso se misurato rispetto a un'unità appropriata, come l'utente o la transazione. Gli obiettivi ti aiutano ad avvalorare un progetto più ampio di sostenibilità che coinvolge la tua azienda o la tua organizzazione, a identificare le regressioni e a dare la priorità a quelle aree che offrono un maggiore potenziale di miglioramento.
- **Ottimizza l'utilizzo:** Dimensiona correttamente i carichi di lavoro e implementa un progetto efficiente in grado di garantire un utilizzo elevato e ottimizzare l'efficienza energetica dell'hardware sottostante. Due host in esecuzione con una percentuale di utilizzo pari al 30% sono meno efficienti di un host in esecuzione al 60%, se consideriamo il consumo di base per host. Allo stesso tempo, elimina o riduci le risorse, le elaborazioni e le archiviazioni inattive per ridurre l'energia totale richiesta per alimentare il tuo carico di lavoro.
- **Anticipa e adotta offerte hardware e software nuove e più efficienti:** Promuovi le migliori a monte di partner e fornitori finalizzate a ridurre l'impatto dei carichi di lavoro cloud. Monitora costantemente il mercato e valuta nuove offerte hardware e software più efficienti. Adotta la flessibilità nei tuoi progetti per consentire una rapida adozione di tecnologie nuove ed efficienti.
- **Utilizza servizi gestiti:** La condivisione dei servizi con un'ampia base clienti consente di ottimizzare l'uso delle risorse e ridurre al tempo stesso l'infrastruttura necessaria per supportare i carichi di lavoro nel cloud. I clienti possono ad esempio condividere l'impatto di componenti comuni di data center, come reti ed energia, migrando i carichi di lavoro su Cloud AWS e adottando servizi gestiti, come AWS Fargate per i container serverless, in cui AWS opera su vasta scala ed è responsabile

della loro efficienza operativa. Utilizza i servizi gestiti per contribuire alla riduzione dell'impatto, trasferendo automaticamente dati con accesso poco frequente all'archiviazione dei dati inattivi con le configurazioni di Amazon S3 Lifecycle o di Amazon EC2 Auto Scaling per adeguare le capacità alla domanda.

- Riduci l'impatto a valle dei carichi di lavoro cloud: Diminuisce la quantità di energia o di risorse richieste per l'utilizzo dei tuoi servizi. Riduci o elimina la necessità di eseguire upgrade dei dispositivi per consentire ai clienti di usare i tuoi servizi. Esegui test usando device farm per analizzare l'impatto atteso e conduci altri test con i clienti per capire l'impatto reale derivante dall'uso dei tuoi servizi.

Processo di miglioramento

Migliorare l'architettura significa capire quali sono le risorse a tua disposizione e cosa puoi fare per migliorare il tuo approccio, selezionando obiettivi di miglioramento, testando i risultati ottenuti, adottando le miglirie più riuscite, quantificando il successo ottenuto e condividendo quanto appreso in modo che possa essere replicato da altri e ripetendo quindi il ciclo.

Gli obiettivi dei tuoi miglioramenti possono essere:

- Eliminare sprechi, scarso utilizzo o risorse inattive o inutilizzate
- Massimizzare il valore di risorse utilizzate

Note

Usa tutte le risorse di cui esegui il provisioning e completa la stessa attività con il minimo di risorse possibile.

In una fase iniziale del processo di ottimizzazione, concentrati sulle aree con sprechi o utilizzo ridotto per poi passare a ottimizzazioni maggiormente finalizzate in linea con il tuo carico di lavoro specifico.

Monitora le variazioni d'uso delle risorse nel tempo. Identifica dove le modifiche accumulate nel tempo causano aumenti significativi e inefficienze nell'uso delle risorse. Definisci le miglirie necessarie per contenere i cambiamenti in termini di consumi e implementa gli interventi più prioritari.

I passaggi seguenti sono stati ideati per generare un processo iterativo che valuta, assegna le priorità, testa e implementa miglirie focalizzate sulla sostenibilità per carichi di lavoro cloud.

1. Identificazione degli obiettivi in termini di miglioramento: esamina i carichi di lavoro rispetto alle best practice di sostenibilità indicate in questo documento e identifica gli obiettivi in termini di miglioramento.
2. Valutazione di miglioramenti specifici: Valuta le modifiche specifiche per identificare i miglioramenti potenziali, i costi previsti e i rischi aziendali.
3. Assegnazione di priorità e pianificazione dei miglioramenti: assegna le priorità alle modifiche che offrono i miglioramenti più sostanziali al costo e con i rischi più bassi e definisci un piano per test e implementazione.

4. Test e convalida dei miglioramenti: implementa le modifiche negli ambienti di test per convalidare il loro potenziale in termini di miglioramenti.
5. Implementazione delle modifiche in produzione: implementa le modifiche negli ambienti di produzione.
6. Misurazione dei risultati e replica dei successi: cerca le opportunità per replicare i successi nei carichi di lavoro e annullare le modifiche con risultati inaccettabili.

Scenario di esempio

Lo scenario di esempio seguente viene utilizzato più avanti in questo documento per illustrare ogni fase del processo di miglioramento.

La tua azienda ha un carico di lavoro che esegue manipolazioni complesse delle immagini su istanze Amazon EC2 e archivia i file originali e modificati per l'accesso degli utenti. Le attività di elaborazione sono operazioni intensive per la CPU e i file di output sono di grandi dimensioni.

Identifica gli obiettivi in termini di miglioramento

Comprendi le best practice che ti aiutano a raggiungere gli obiettivi di sostenibilità. Puoi trovare descrizioni dettagliate di queste [best practice](#) e suggerimenti per miglorie più avanti in questo documento.

Esamina i carichi di lavoro e le risorse utilizzate. Identifica gli hot spot come grandi implementazioni e risorse utilizzate di frequente. Valuta questi hot spot per avere la possibilità di migliorare l'utilizzo effettivo delle risorse e ridurre le risorse totali necessarie per ottenere i risultati aziendali auspicati.

Esamina il carico di lavoro rispetto alle best practice e identifica i candidati per il miglioramento.

Applicando questo passaggio a [Scenario di esempio](#), identifichi le seguenti best practice come obiettivi probabili per il miglioramento:

- Utilizza la quantità minima di hardware per soddisfare le tue esigenze
- Utilizza tecnologie che supportano al meglio l'accesso ai dati e i modelli di storage

Risorse

- [Optimizing your AWS Infrastructure for Sustainability, Part I: Compute \(Ottimizzazione dell'infrastruttura AWS per la sostenibilità, Parte I: Calcolo\)](#)

- [Optimizing your AWS Infrastructure for Sustainability, Part II: Storage \(Ottimizzazione dell'infrastruttura AWS per la sostenibilità, Parte II: Archiviazione\)](#)
- [Optimizing your AWS Infrastructure for Sustainability, Part III: Networking \(Ottimizzazione dell'infrastruttura AWS per la sostenibilità, Parte III: Reti\)](#)

Valuta miglioramenti specifici

Scopri le risorse fornite dal tuo carico di lavoro per completare un'unità di lavoro. Valuta le migliori potenziali e stima il loro impatto previsto, il costo dell'implementazione e i rischi associati.

Per misurare le migliori nel tempo, analizza innanzitutto cosa hai fornito in AWS e come queste risorse sono state utilizzate.

Comincia da una panoramica completa dell'utilizzo di AWS e usa Report di costi e utilizzo AWS per identificare gli hot spot. Usa questo [codice di esempio AWS](#) per aiutarti a esaminare e analizzare il report con l'aiuto di Amazon Athena.

Metriche proxy

Quando valuti modifiche specifiche devi anche valutare quali metriche sono in grado di quantificare meglio l'effetto di tale modifica sulla risorsa associata. Queste metriche sono chiamate metriche proxy. Seleziona le metriche proxy che meglio rappresentano il tipo di miglioramento che stai valutando e le risorse coinvolte nel miglioramento. Queste metriche possono evolvere con il tempo.

Le risorse fornite per supportare il carico di lavoro includono elaborazione, archiviazione e risorse di rete. Valuta le risorse fornite usando le metriche proxy per vedere come queste risorse vengono utilizzate.

Usa le metriche proxy per misurare le risorse fornite per raggiungere i risultati aziendali.

Risorsa	Metriche proxy di esempio	Obiettivi di miglioramento
Calcolo	Minuti vCPU	Ottimizza l'utilizzo delle risorse fornite
Storage	GB forniti	Riduci il totale fornito
Network	GB o pacchetti trasferiti	Riduci il totale trasferito e la distanza trasferita

Metriche aziendali

Seleziona metriche aziendali per quantificare il raggiungimento degli obiettivi di business. Le metriche di business devono rispecchiare il valore fornito dal carico di lavoro, ad esempio, il numero di utenti attivi simultaneamente, chiamate API servite o il numero di transazioni completate. Queste metriche possono evolvere con il tempo. Fai attenzione quando valuti metriche aziendali basate su aspetti finanziari, poiché l'incoerenza del valore delle transazioni rende nulli i paragoni.

Indicatori chiave di prestazione

Usando la formula seguente, dividi le risorse fornite in base ai risultati di business ottenuti per stabilire le risorse fornite per unità di lavoro.

$$\text{Risorse fornite per unità di lavoro} = \frac{\text{Metrica proxy per la risorsa fornita}}{\text{Metrica di business per il risultato}}$$

Formula KPI

Usa le risorse per unità di lavoro come i KPI. Stabilisci previsioni in base alle risorse fornite come base per i paragoni.

Risorsa	KPI di esempio	Obiettivi di miglioramento
Calcolo	Minuti vCPU per transazione	Ottimizza l'utilizzo delle risorse fornite
Storage	GB per transazione	Riduci il totale fornito
Network	GB trasferiti per transazione o pacchetti trasferiti per transazione	Riduci il totale trasferito e la distanza trasferita

Stima del miglioramento

Stima il miglioramento come riduzione quantitativa delle risorse fornite (come indicato dalle metriche proxy) e la percentuale di modifica rispetto alle risorse principali fornite per unità di lavoro.

Risorsa	KPI di esempio	Obiettivi di miglioramento
Calcolo	% di riduzione dei minuti di vCPU per transazione	Ottimizza l'utilizzo
Storage	% di riduzione di GB per transazione	Riduci il totale fornito
Network	% di riduzione dei GB trasferiti per transazione o pacchetti trasferiti per transazione	Riduci il totale trasferito e la distanza trasferita

Stima dei miglioramenti

Valuta i miglioramenti potenziali rispetto ai vantaggi netti previsti. Valuta i tempi, i costi e il livello di impegno per l'implementazione e la manutenzione, nonché i rischi aziendali per gli impatti non previsti.

I miglioramenti target spesso rappresentano compromessi tra i tipi di risorse utilizzati. Ad esempio, per ridurre l'uso di elaborazioni, puoi archiviare un risultato, limitare i dati trasferiti o elaborare i dati prima di inviare il risultato a un cliente. Questi [compromessi](#) sono discussi più avanti in maggior dettaglio.

Nella fase di valutazione dei rischi per il tuo carico di lavoro includi requisiti non funzionali, tra cui sicurezza, affidabilità, efficienza delle performance, ottimizzazione dei costi e impatto dei miglioramenti sulla tua capacità di gestire il carico di lavoro.

Applicando questo passaggio a [Scenario di esempio](#), valuta i miglioramenti target con i seguenti risultati:

Best practice	Miglioramento target	Potenziale	Costo	Rischio
Utilizza la quantità minima di hardware per	Implementa dimensionamenti predittivi per	Medium	Bassa	Bassa

Best practice	Miglioramento target	Potenziale	Costo	Rischio
soddisfare le tue esigenze	ridurre i periodi di scarso utilizzo			
Utilizza tecnologie che supportano al meglio l'accesso ai dati e i modelli di storage	Implementa meccanismi di compressione più efficaci per ridurre l'archiviazione totale e il tempo necessario per completarla	Alta	Bassa	Bassa

Implementando la pianificazione predittiva è possibile ridurre le ore di vCPU consumate da istanze sottoutilizzate o non utilizzate che forniscono vantaggi moderati rispetto a meccanismi di dimensionamento esistenti con una riduzione stimata dell'11% delle risorse utilizzate. I costi previsti sono bassi e includono la configurazione delle risorse cloud e la gestione di un dimensionamento predittivo per Amazon EC2 Auto Scaling. Il rischio limita le performance quando la riduzione viene applicata come reazione in risposta a previsioni che superano la domanda.

L'implementazione di una compressione più efficace avrà un impatto significativo con riduzione significativa della dimensione dei file in tutte le immagini originali e manipolate, con una riduzione stimata del 25% in termini di requisiti di archiviazione in produzione. L'implementazione di un nuovo algoritmo è una sostituzione che richiede poco impegno e non comporta molti rischi.

Assegna le priorità e pianifica i miglioramenti

Assegna le priorità ai miglioramenti identificati in base all'impatto previsto di dimensioni maggiori con i costi più bassi e un livello di rischio accettabile.

Decidi su quali miglioramenti concentrarti inizialmente e includili nella tua pianificazione delle risorse e nella roadmap di sviluppo.

Applicando questo passaggio a [Scenario di esempio](#), assegna la priorità ai miglioramenti target come segue:

Priorità	miglioramento KPI	Potenziale	correnti	Rischio
1	Implementa meccanismi di compressione più efficaci	Alta	Bassa	Bassa
2	Implementa il dimensionamento predittivo	Medium	Bassa	Bassa

L'aggiornamento della compressione dei file offre un potenziale elevato, un costo ridotto e rischi contenuti, rendendolo così un'operazione ad alto valore per la tua azienda e una priorità rispetto all'implementazione di un dimensionamento predittivo. Stabilisci che l'implementazione del dimensionamento predittivo con il suo potenziale medio di impatto, il costo contenuto e il rischio basso debba essere il miglioramento con priorità più alta una volta completata la compressione dei file.

Incarichi un membro del team di implementare la compressione dei file ottimizzata e aggiungi il dimensionamento predittivo al backlog.

Testa e convalida i miglioramenti

Esegui piccoli test con investimenti ridotti per ridurre il rischio di un impegno su larga scala.

Implementa una copia rappresentativa del tuo carico di lavoro nell'ambiente di test per limitare i costi e i rischi legati a test e convalide. Esegui un set predefinito di transazioni di test, misura le risorse fornite e stabilisci le risorse usate per unità di lavoro per definire una base per il test.

Implementa il miglioramento target nell'ambiente di test e ripeti il test usando la stessa metodologia nelle stesse condizioni. Quindi, misura le risorse fornite e le risorse utilizzate per unità di lavoro con il tuo miglioramento attivo.

Calcola la percentuale di modifica rispetto alle risorse di base fornite per unità di lavoro e stabilisci la riduzione quantitativa attesa in termini di risorse fornite nell'ambiente di produzione. Confronta questi valori con i valori previsti. Stabilisci se il risultato raggiunge un livello accettabile di

miglioramento. Valuta se i compromessi nelle risorse aggiuntive utilizzate rendono il beneficio netto del miglioramento inaccettabile.

Stabilisci se il miglioramento è un successo e se le risorse devono essere investite nell'implementazione della modifica in produzione. Se a questo punto la modifica viene valutata come un insuccesso, reindirizza le tue risorse per testare e convalidare il tuo target successivo e continuare con il tuo ciclo di miglioramenti.

% di riduzione delle risorse fornite per unità di lavoro	Riduzione quantitativa delle risorse fornite	Operazione
Soddisfa le aspettative	Soddisfa le aspettative	Procedere con il miglioramento
Non soddisfa le aspettative	Soddisfa le aspettative	Procedere con il miglioramento
Soddisfa le aspettative	Non soddisfa le aspettative	Valutare un miglioramento alternativo
Non soddisfa le aspettative	Non soddisfa le aspettative	Valutare un miglioramento alternativo

Applicando questo passaggio a [Scenario di esempio](#), esegui i test per convalidare il successo.

Dopo aver eseguito i test sull'algoritmo di compressione ottimizzato, la riduzione in percentuale delle risorse fornite per unità di lavoro (l'archiviazione richiesta per l'immagine originale e modificata) soddisfa le aspettative con una riduzione media del 30% nell'archiviazione fornita e un carico di calcolo aumentato in quantità irrilevante.

Stabilisci che le risorse di calcolo aggiuntive richieste per applicare l'algoritmo di compressione ottimizzato ai file esistenti in produzione sono irrilevanti rispetto alla riduzione dell'archiviazione ottenuta. Confermi il risultato positivo della riduzione quantitativa in termini di risorse richieste (TB di archiviazione) e il miglioramento viene approvato per l'implementazione in produzione.

Implementazione delle modifiche in produzione

Implementa in produzione miglioramenti testati, convalidati e approvati. Ottimizza usando implementazioni limitate, conferma il funzionamento del tuo carico di lavoro, testa la riduzione reale delle risorse fornite e delle risorse usate per unità di lavoro nell'implementazione limitata e verifica le conseguenze non intenzionali della modifica. Procedi con le implementazioni complete dopo il superamento dei test.

Elimina le modifiche se il test ha esito negativo o se si verificano conseguenze non previste e inaccettabili dovute alla modifica.

Applicando questo passaggio a [Scenario di esempio](#), intraprendi le seguenti azioni.

Realizza le modifiche in produzione tramite una implementazione limitata con una metodologia di implementazione blu/verde. I test di funzionamento rispetto alle nuove istanze implementate hanno avuto esito positivo. Hai registrato una riduzione in media del 26% in termini di archiviazione fornita per file di immagini manipolate e originali. Non vedi evidenza di un aumento nelle risorse di calcolo con la compressione dei nuovi file.

Noti una diminuzione non prevista del tempo trascorso per comprimere i file di immagini e lo attribuisce al codice altamente ottimizzato per il nuovo algoritmo di compressione.

Procedi con l'implementazione completa della nuova versione.

Misurazione dei risultati e replica dei successi

Misura i risultati e replica i successi nei seguenti modi:

- Misura il miglioramento iniziale registrato dalle risorse fornite per unità di lavoro e la diminuzione quantitativa delle risorse fornite.
- Confronta le stime iniziali e i risultati dei test con le misurazioni in produzione. Identifica i fattori che possono aver contribuito alle differenze e aggiorna le tue stime e le metodologie di test, se necessario.
- Stabilisci il successo, e il grado di successo, e condividi i risultati con le parti interessate.
- Se devi annullare le modifiche in seguito al risultato negativo dei test o a conseguenze negative non previste della modifica, identifica i fattori che hanno contribuito a questo risultato. Itera laddove possibile o valuta nuovi approcci per ottenere gli obiettivi della modifica.

- Prendi ciò che hai appreso, stabilisci gli standard e applica miglioramenti positivi ad altri sistemi che possono trarne ugualmente vantaggio. Acquisisci e condividi la tua metodologia, gli artefatti relativi e i benefici netti con team e organizzazioni in modo che altri possano adottare questo standard e replicare il successo da te ottenuto.
- Monitora le risorse fornite per unità di lavoro e tieni traccia delle modifiche e dell'impatto totale nel tempo. Le modifiche al carico di lavoro o a come i clienti usano il carico di lavoro può avere un impatto sull'efficacia del tuo miglioramento. Valuta nuovamente le opportunità di miglioramento se noti diminuzioni significative a breve termine dell'efficacia della modifica o un'accumulata riduzione dell'efficacia nel tempo.
- Quantifica il beneficio netto che deriva nel tempo dal tuo miglioramento (inclusi i vantaggi ricevuti da altri team che hanno applicato il tuo miglioramento, se disponibile) per dimostrare il ritorno sugli investimenti derivante da attività di questo tipo.

Applicando questo passaggio a [Scenario di esempio](#), misuri i risultati seguenti.

Il carico di lavoro mostra un miglioramento iniziale con una riduzione del 23% dei requisiti di archiviazione dopo l'implementazione e l'applicazione del nuovo algoritmo di compressione ai file di immagine esistenti.

Il valore misurato è largamente in linea con le stime iniziali (25%) e la differenza significativa rispetto al test (30%) dipende dal fatto che i file di immagine usati nel test non sono rappresentativi dei file di immagine presenti in produzione. Modifichi l'immagine di test per rappresentare in modo più appropriato le immagini in produzione.

Il miglioramento è considerato un successo sotto ogni aspetto. La riduzione totale dell'archiviazione fornita è inferiore del 2% rispetto alla stima del 25%, ma il 23% è comunque un miglioramento significativo in termini di impatto sulla sostenibilità ed è corredato da un risparmio equivalente sui costi.

Le uniche conseguenze non previste della modifica sono la riduzione utile del tempo trascorso per eseguire la compressione e l'equivalente riduzione del vCPU usato. Queste miglorie sono attribuite al codice altamente ottimizzato.

Definisci un progetto open-source interno in cui condividere il codice, gli artefatti associati, le linee guida su come implementare la modifica e i risultati che ne derivano. Il progetto interno open-source semplifica l'adozione del codice da parte dei team in tutti i casi d'uso di archiviazione persistente dei file. I team adottano il miglioramento come standard. I benefici secondari del progetto open-source

interno sono i vantaggi che derivano dai miglioramenti alla soluzione di cui tutti possono usufruire adottandola e qualsiasi persona può contribuire ai miglioramenti al progetto.

Pubblica il successo conseguito e condividi il progetto open-source all'interno dell'organizzazione. Ogni team che adotta la soluzione replica il vantaggio con un investimento minimo e aggiunge il beneficio netto ricevuto dall'investimento. Pubblichiamo questi dati come una storia di successo continua nel tempo.

Continui a monitorare l'impatto del miglioramento nel tempo ed effettui le modifiche al progetto open-source interno in base alle esigenze.

Sostenibilità come requisito non funzionale

Aggiungere la sostenibilità al proprio elenco di requisiti aziendali può consentire di ottenere risultati più vantaggiosi dal punto di vista dei costi. Se ti focalizzi su come ottenere maggior valore dalle risorse usandone di meno, ottieni direttamente una riduzione dei costi su AWS, poiché paghi solo ciò che usi.

Soddisfare gli obiettivi di sostenibilità non comporta necessariamente compromessi in altre metriche tradizionali, come i tempi di attività, la disponibilità o i tempi di risposta. Tuttavia, spesso è possibile ottenere vantaggi significativi in termini di sostenibilità senza alcun impatto misurabile sui livelli di servizio. Laddove sono richiesti piccoli compromessi, il miglioramento in termini di sostenibilità così ottenuto può superare il cambiamento in termini di qualità del servizio.

Incoraggia i membri del tuo team a sperimentare continuamente migliorie in termini di sostenibilità durante la fase di sviluppo di requisiti funzionali. I team devono anche integrare metriche relative ai proxy nella fase di definizione degli obiettivi per avere la certezza di valutare l'intensità delle risorse durante lo sviluppo dei carichi di lavoro.

Ecco alcuni esempi dei compromessi che possono ridurre le risorse cloud che usi:

Adegua la qualità del risultato: Puoi scambiare la Quality of Results (QoR) in favore di una riduzione dell'intensità del carico di lavoro con un calcolo approssimativo. La pratica dell'elaborazione approssimativa cerca opportunità per sfruttare il divario tra ciò che è assolutamente necessario e ciò che effettivamente produci. Ad esempio, se i dati vengono inseriti in una struttura di dati impostata , puoi eliminare l'operatore ORDER BY in SQL per rimuovere elaborazioni superflue, ridurre l'uso di risorse e offrire, al tempo stesso, una risposta accettabile.

Adegua i tempi di risposta: Una risposta con un tempo di risposta più lento può favorire il risparmio di anidride carbonica, riducendo al minimo il sovraccarico condiviso. L'elaborazione di attività temporanee ad hoc può comportare un sovraccarico in fase di avvio. Raggruppa le attività ed elaborale in batch, invece di pagare per il sovraccarico ogni volta che arriva un'attività. L'elaborazione in batch implica l'aumento del tempo di risposta in cambio di una riduzione del sovraccarico condiviso della creazione di un'istanza, del download del codice sorgente e dell'esecuzione del processo.

Adegua la disponibilità: AWS consente di aggiungere facilmente ridondanza e di raggiungere target di elevata disponibilità con pochi e semplici clic. Puoi aumentare la ridondanza con tecniche come la stabilità statica, eseguendo il provisioning di risorse inattive che comportano sempre una riduzione dell'utilizzo. Valuta le esigenze dell'azienda quando definisci gli obiettivi. Compromessi relativamente

minori in termini di disponibilità possono portare a miglioramenti molto più significativi in termini di utilizzo. Ad esempio, un modello di architettura di stabilità statica prevede il provisioning della capacità di failover inattiva per assorbire immediatamente il carico dopo un guasto di un componente. Allentare i requisiti di disponibilità può eliminare la necessità di capacità online inattiva consentendo all'automazione di distribuire risorse sostitutive. L'aggiunta di capacità di failover on demand consente un utilizzo complessivo più elevato senza alcun impatto sull'attività durante le normali operazioni e offre il vantaggio secondario di ridurre i costi.

Best practice per la sostenibilità nel cloud

Ottimizza il posizionamento dei carichi di lavoro e l'architettura per domanda, software, dati, hardware e processi al fine di aumentare l'efficienza energetica. Ognuna di queste aree offre l'opportunità di adottare best practice per ridurre l'impatto dei carichi di lavoro cloud in termini di sostenibilità, ottimizzando l'utilizzo e riducendo al minimo gli sprechi e la quantità totale di risorse implementate e alimentate per supportare il carico di lavoro.

Argomenti

- [Selezione delle regioni](#)
- [Allineamento alla domanda](#)
- [Software e architettura](#)
- [Gestione dati](#)
- [Hardware e servizi](#)
- [Processo e cultura](#)

Selezione delle regioni

La scelta della Regione per il carico di lavoro influisce in modo significativo sui suoi KPI, tra cui prestazioni, costi e impatto ambientale. Per migliorare in modo efficace questi KPI, devi scegliere le regioni per i carichi di lavoro in base alle esigenze aziendali e agli obiettivi di sostenibilità.

Best practice

- [SUS01-BP01 Scelta della Regione in base alle esigenze aziendali e agli obiettivi di sostenibilità.](#)

SUS01-BP01 Scelta della Regione in base alle esigenze aziendali e agli obiettivi di sostenibilità.

Scegli la Regione del tuo carico di lavoro in base alle esigenze aziendali e agli obiettivi di sostenibilità per ottimizzare i suoi KPI, tra cui prestazioni, costi e impatto ambientale.

Anti-pattern comuni:

- Selezione della Regione del carico di lavoro in base alla propria collocazione.
- Consolidamento di tutte le risorse del carico di lavoro in un'unica posizione geografica.

Vantaggi dell'adozione di questa best practice: la collocazione di un carico di lavoro in prossimità di progetti legati alla generazione di energia rinnovabile di Amazon o in Regioni con un'intensità ridotta di emissione di anidride carbonica nota può contribuire a ridurre il suo impatto ambientale.

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

Guida all'implementazione

Il Cloud AWS è una rete in costante espansione di Regioni e punti di presenza (PoP), con un'infrastruttura di rete globale che li collega tra loro. La scelta della Regione per il carico di lavoro influisce in modo significativo sui suoi KPI, tra cui prestazioni, costi e impatto ambientale. Per migliorare efficacemente questi KPI, è necessario scegliere le Regioni per il proprio carico di lavoro in base alle esigenze aziendali e agli obiettivi di sostenibilità.

Passaggi dell'implementazione

- Segui questi passaggi per valutare e selezionare le potenziali Regioni per il tuo carico di lavoro in base ai requisiti aziendali, tra cui la conformità, le funzionalità disponibili, il costo e la latenza.
 - Conferma che queste Regioni siano conformi in base alle normative locali richieste.
 - Utilizza gli [elenchi dei servizi regionali di AWS](#) per verificare che le Regioni dispongano dei servizi e delle funzionalità necessarie per gestire il tuo carico di lavoro.
 - Calcola il costo del carico di lavoro su ogni Regione utilizzando [AWS Pricing Calculator](#).
 - Valuta la latenza di rete tra le sedi degli utenti finali e ogni Regione AWS.
- Scegli le Regioni in prossimità dei progetti di generazione di energia rinnovabile di Amazon e le Regioni in cui la griglia presenta un'intensità di emissione di anidride carbonica nota inferiore a quella di altre sedi (o Regioni).
 - Identifica le tue principali linee guida sulla sostenibilità per tracciare e confrontare le emissioni di anidride carbonica anno per anno sulla base del [Greenhouse Gas Protocol](#) (metodi basati sul mercato e sulla localizzazione).
 - Scegli la Regione in base al metodo utilizzato per monitorare le emissioni di anidride carbonica. Per maggiori dettagli sulla scelta di una Regione in base alle linee guida sulla sostenibilità, consulta [Come selezionare una Regione per il carico di lavoro in base agli obiettivi di sostenibilità](#).

Risorse

Documenti correlati:

- [Comprendere le stime delle emissioni di anidride carbonica](#)
- [Amazon Around the Globe](#)
- [Metodologia delle energie rinnovabili](#)
- [Quali elementi valutare quando si seleziona una Regione per i propri carichi di lavoro](#)

Video correlati:

- [AWS re:Invent 2023 - Sustainability innovation in AWS Global Infrastructure](#)
- [AWS re:Invent 2023 - Sustainable architecture: Past, present, and future](#)
- [AWS re:Invent 2022 - Delivering sustainable, high-performing architectures](#)
- [AWS re:Invent 2022 - Architecting sustainably and reducing your AWS carbon footprint](#)
- [AWS re:Invent 2022 - Sustainability in AWS global infrastructure](#)

Allineamento alla domanda

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

Best practice

- [SUS02-BP01 Scala dinamicamente l'infrastruttura dei carichi di lavoro](#)
- [SUS02-BP02 Allineamento degli SLA agli obiettivi di sostenibilità](#)
- [SUS02-BP03 Interruzione della creazione e della manutenzione di risorse inutilizzate](#)
- [SUS02-BP04 Ottimizzazione del posizionamento geografico dei carichi di lavoro in base ai requisiti di rete](#)
- [SUS02-BP05 Ottimizzazione delle risorse dei membri del team in base alle attività eseguite](#)
- [SUS02-BP06 Implementazione del buffering o della limitazione \(della larghezza di banda della rete\) per ridurre la curva della domanda](#)

SUS02-BP01 Scala dinamicamente l'infrastruttura dei carichi di lavoro

Usa l'elasticità del cloud e dimensiona la tua infrastruttura in modo dinamico per rispondere alla richiesta di fornitura di risorse cloud ed evitare capacità sovra-assegnate nel tuo carico di lavoro.

Anti-pattern comuni:

- Mancato dimensionamento dell'infrastruttura in base al carico degli utenti.
- Costante dimensionamento manuale dell'infrastruttura.
- Dopo un evento di dimensionamento, lasci una capacità aumentata anziché ridurre il dimensionamento.

vantaggi derivanti dall'applicazione di questa best practice: la configurazione e il test dell'elasticità dei carichi di lavoro aiuta ad abbinare correttamente richiesta e fornitura di risorse cloud e a evitare capacità sovra-assegnate. Puoi sfruttare i vantaggi dell'elasticità nel cloud per dimensionare automaticamente la capacità durante e dopo i picchi di richiesta ed essere sicuro di utilizzare solo il numero esatto di risorse necessario per soddisfare le esigenze aziendali.

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

Guida all'implementazione

Il cloud offre la flessibilità necessaria per espandere o ridurre le risorse in modo dinamico attraverso una serie di meccanismi per soddisfare i cambiamenti della domanda. La corrispondenza ottimale tra offerta e domanda consente l'impatto ambientale più basso per un carico di lavoro.

La domanda può essere fissa o variabile e richiede parametri e automazione, allo scopo di garantire che la gestione non diventi particolarmente onerosa. Le applicazioni possono essere dimensionate verticalmente (verso l'alto o verso il basso) modificando la dimensione dell'istanza, orizzontalmente (aumentando o diminuendo) modificando il numero di istanze o tramite una combinazione delle due opzioni.

Puoi adottare varie strategie di approccio per associare l'offerta di risorse alla domanda.

- Approccio di monitoraggio del target: monitora il parametro di dimensionamento e aumenta o diminuisci automaticamente la capacità in base alle esigenze.
- Dimensionamento predittivo: dimensiona l'anticipazione di tendenze giornaliere e settimanali.
- Approccio basato sulla pianificazione: imposta la tua pianificazione di dimensionamento in base a modifiche di carico prevedibili.

- Dimensionamento dei servizi: scegli i servizi (come il serverless) che usano il dimensionamento in modo nativo per impostazione predefinita o che forniscono il dimensionamento automatico come funzionalità.

Identifica i periodi di utilizzo assente o ridotto e dimensiona le risorse per evitare capacità in eccesso e migliorare il livello di efficienza.

Passaggi dell'implementazione

- L'elasticità corrisponde all'offerta di risorse disponibili rispetto alla relativa domanda. Istanze, container e funzioni offrono meccanismi di elasticità, sia insieme al dimensionamento automatico sia come funzionalità del servizio. AWS offre una gamma di meccanismi di dimensionamento automatico per avere la certezza che i carichi di lavoro possano essere ridotti facilmente e velocemente nei periodi di basso carico di utenti. Ecco alcuni esempi di meccanismi di dimensionamento automatico:

Auto scaling mechanism	Where to use
Amazon EC2 Auto Scaling	Usalo per verificare che sia disponibile il numero corretto di istanze Amazon EC2 per gestire il carico degli utenti dell'applicazione.
Application Auto Scaling	Usalo per dimensionare automaticamente le risorse per servizi AWS diversi da Amazon EC2, ad esempio funzioni Lambda o servizi Amazon Elastic Container Service (Amazon ECS).
Kubernetes Cluster Autoscaler	Usalo per dimensionare automaticamente i cluster Kubernetes su AWS.

- Si parla spesso di dimensionamento con servizi di elaborazione come le istanze Amazon EC2 o le funzioni AWS Lambda. Considera la configurazione di servizi non di elaborazione come unità di capacità di lettura e scrittura [Amazon DynamoDB](#) o partizioni [Amazon Kinesis Data Streams](#) per rispondere alle richieste.
- Verifica che le metriche per il dimensionamento verticale o orizzontale siano convalidate in base al tipo di carico di lavoro implementato. Se distribuisce un'applicazione di transcodifica video, è previsto il 100% di utilizzo della CPU e non deve essere il parametro principale. Se

necessario, puoi usare una [metrica personalizzata](#) (come l'uso della memoria) per la tua politica di dimensionamento. Per scegliere la metrica corretta, consulta le linee guida seguenti per Amazon EC2:

- La metrica deve essere una metrica di utilizzo valida e descrivere il livello di impiego di un'istanza.
- Il valore della metrica deve aumentare o diminuire proporzionalmente in base al numero di istanze nel gruppo con Auto Scaling.
- Usa il [dimensionamento dinamico](#) invece del [dimensionamento manuale](#) per il tuo gruppo Auto Scaling. Ti consigliamo anche di usare [politiche di dimensionamento del monitoraggio degli obiettivi](#) nel tuo dimensionamento dinamico.
- Verifica che le implementazioni dei carichi di lavoro siano in grado di gestire eventi di dimensionamento orizzontale. Crea scenari di test per eventi di dimensionamento orizzontale per verificare che il carico di lavoro si comporti secondo le aspettative e che non incida sull'esperienza utente (come nel caso della perdita di sessioni permanenti). Puoi usare la [Cronologia delle attività](#) per verificare un'attività di dimensionamento per un gruppo Auto Scaling.
- Analizza il tuo carico di lavoro per individuare modelli prevedibili e dimensionare le tue risorse in modo proattivo, anticipando variazioni nella domanda previste e pianificate. Con il dimensionamento predittivo puoi eliminare la necessità di offrire capacità in eccedenza. Per maggiori dettagli consulta [Dimensionamento predittivo con Amazon EC2 Auto Scaling](#).

Risorse

Documenti correlati:

- [Nozioni di base su Amazon EC2 Auto Scaling](#)
- [Dimensionamento predittivo per EC2, alimentato dal machine learning](#)
- [Analizza il comportamento degli utenti tramite Amazon OpenSearch Service, Amazon Data Firehose e Kibana](#)
- [Che cos'è Amazon CloudWatch?](#)
- [Monitoraggio del carico del database con Performance Insights su Amazon RDS](#)
- [Introduzione al supporto nativo per il dimensionamento predittivo con Amazon EC2 Auto Scaling](#)
- [Introducing Karpenter - An Open-Source, High-Performance Kubernetes Cluster Autoscaler \(Introduzione a Karpenter - Kubernetes Cluster Autoscaler, uno strumento open source a elevate prestazioni\)](#)

- [Deep Dive su Amazon ECS Cluster Auto Scaling](#)

Video correlati:

- [AWS re:Invent 2023 - Scaling on AWS for the first 10 million users](#)
- [AWS re:Invent 2023 - Sustainable architecture: Past, present, and future](#)
- [AWS re:Invent 2022 - Build a cost-, energy-, and resource-efficient compute environment](#)
- [AWS re:Invent 2022 - Scaling containers from one user to millions](#)
- [AWS re:Invent 2023 - Scaling FM inference to hundreds of models with Amazon SageMaker](#)
- [AWS re:Invent 2023 - Harness the power of Karpenter to scale, optimize & upgrade Kubernetes](#)

Esempi correlati:

- [Dimensionamento automatico](#)

SUS02-BP02 Allineamento degli SLA agli obiettivi di sostenibilità

Rivedi e ottimizza gli Accordi sul livello di servizio (SLA) del carico di lavoro in base ai tuoi obiettivi di sostenibilità per ridurre la quantità di risorse richieste per supportare il carico di lavoro e continuare a soddisfare le esigenze aziendali.

Anti-pattern comuni:

- Gli SLA dei carichi di lavoro sono sconosciuti o ambigui.
- Definisci il tuo SLA solo per disponibilità e performance.
- Usi lo stesso modello di progettazione (come l'architettura multi-AZ) per tutti i carichi di lavoro.

Vantaggi dell'adozione di questa best practice: allineare gli SLA agli obiettivi di sostenibilità porta a un utilizzo ottimale delle risorse e, al contempo, a una conciliazione con le esigenze aziendali.

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

Guida all'implementazione

Gli SLA definiscono il livello di servizio che ci si aspetta da un carico di lavoro cloud, come ad esempio i tempi di risposta, la disponibilità e la conservazione dei dati. Essi influenzano l'architettura,

l'utilizzo delle risorse e l'impatto ambientale di un carico di lavoro nel cloud. Con cadenza regolare, rivedi gli SLA e accetta dei compromessi che riducano l'utilizzo di risorse in modo significativo in cambio di una diminuzione accettabile dei livelli di servizio.

Passaggi dell'implementazione

- Comprendi gli obiettivi di sostenibilità: individua nella tua organizzazione gli obiettivi di sostenibilità, come la riduzione delle emissioni di carbonio o il miglioramento dell'utilizzo delle risorse.
- Rivedi gli SLA: esamina gli accordi sul livello di servizio (SLA) per valutare se supportano i requisiti aziendali. Se stai superando gli SLA, esegui un'ulteriore revisione.
- Comprendi i compromessi: individua i compromessi tra la complessità del carico di lavoro, ad esempio un elevato volume di utenti simultanei, le prestazioni, ad esempio la latenza, e l'impatto sulla sostenibilità, ad esempio le risorse richieste. In genere, dare la priorità a due fattori va a scapito del terzo.
- Modifica gli SLA: cambia gli SLA accettando compromessi che riducano l'impatto in termini di sostenibilità in modo significativo in cambio di una diminuzione accettabile dei livelli di servizio.
 - Sostenibilità e affidabilità: i carichi di lavoro altamente disponibili tendono a consumare più risorse.
 - Sostenibilità e performance: l'uso di una maggiore quantità di risorse per aumentare le performance potrebbe avere un impatto ambientale più significativo.
 - Sostenibilità e sicurezza: carichi di lavoro con una sicurezza eccessiva potrebbero avere un impatto maggiore sull'ambiente.
- Definisci gli SLA di sostenibilità, se possibile: includi gli SLA di sostenibilità per il tuo carico di lavoro. Ad esempio, determina un livello minimo di utilizzo come SLA di sostenibilità per le tue istanze di calcolo.
- Usa modelli di progettazione efficienti: utilizza modelli di progettazione, come i microservizi su AWS, che danno la priorità a funzioni strategiche per la tua azienda e consentono livelli di servizio inferiori (in tema di obiettivi per tempi di risposta o di ripristino) per funzioni non critiche.
- Comunica e stabilisci le responsabilità: condividi gli SLA con tutte le parti interessate pertinenti, inclusi il team di sviluppo e i tuoi clienti. Utilizza i report per tracciare e monitorare gli SLA. Assegna le responsabilità per raggiungere gli obiettivi di sostenibilità degli SLA.
- Usa incentivi e premi: utilizza incentivi e premi per raggiungere o superare gli SLA in linea con gli obiettivi di sostenibilità.
- Rivedi e itera: revisiona e modifica regolarmente i tuoi SLA per assicurarti che siano allineati agli obiettivi di sostenibilità e prestazioni in costante evoluzione.

Risorse

Documenti correlati:

- [Understand resiliency patterns and trade-offs to architect efficiently in the cloud](#)
- [L'importanza del contratto sul livello di servizi \(SLA\) per i provider SaaS](#)

Video correlati:

- [AWS re:Invent 2023 - Capacity, availability, cost efficiency: Pick three](#)
- [AWS re:Invent 2023 - Sustainable architecture: Past, present, and future](#)
- [AWS re:Invent 2023 - Advanced integration patterns & trade-offs for loosely coupled systems](#)
- [AWS re:Invent 2022 - Delivering sustainable, high-performing architectures](#)
- [AWS re:Invent 2022 - Build a cost-, energy-, and resource-efficient compute environment](#)

SUS02-BP03 Interruzione della creazione e della manutenzione di risorse inutilizzate

Disattiva le risorse non utilizzate nel tuo carico di lavoro per ridurre il numero di risorse cloud richieste per supportare la domanda e per ridurre gli sprechi.

Anti-pattern comuni:

- Non analizzi la tua applicazione per individuare le risorse ridondanti o non più necessarie.
- Non rimuovi le risorse ridondanti o non più necessarie.

Vantaggi dell'adozione di questa best practice: se si eliminano le risorse non utilizzate si libera capacità e si migliora l'efficienza generale del carico di lavoro.

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

Guida all'implementazione

Le capacità inutilizzate consumano risorse cloud come spazio di archiviazione e potenza di elaborazione. Individuando ed eliminando queste risorse, puoi liberare capacità e ottenere un'architettura cloud più efficiente. Analizza le risorse delle applicazioni con regolarità (come report precompilati, set di dati, immagini statiche e modelli di accesso alle risorse) per identificare

ridondanze, sottoutilizzi e obiettivi potenziali di disattivazione. Elimina le risorse ridondanti per ridurre gli sprechi nel tuo carico di lavoro.

Passaggi dell'implementazione

- Redigi l'inventario: esegui l'inventario completo per identificare tutte le risorse del tuo carico di lavoro.
- Analizza l'utilizzo: usa il monitoraggio continuo per identificare le risorse statiche che non sono più necessarie.
- Rimuovi le risorse inutilizzate: sviluppa un piano per rimuovere le risorse che non sono più necessarie.
 - Prima di rimuovere qualsiasi risorsa, valuta l'impatto della rimozione sull'architettura.
 - Analizza le risorse generate in sovrapposizione per rimuovere le elaborazioni ridondanti.
 - Aggiorna le tue applicazioni per smettere di produrre e archiviare risorse che non sono più necessarie.
- Comunica con le terze parti: istruisci le terze parti affinché smettano di produrre e archiviare per tuo conto risorse gestite non più necessarie. Chiedi di consolidare le risorse ridondanti.
- Usa le policy del ciclo di vita: utilizza le policy del ciclo di vita per eliminare automaticamente le risorse inutilizzate.
 - Puoi utilizzare il ciclo di vita Amazon S3 per gestire gli oggetti durante il loro ciclo di vita.
 - Puoi utilizzare Amazon Data Lifecycle Manager per automatizzare la creazione, la conservazione e l'eliminazione degli snapshot Amazon EBS e delle AMI basate su Amazon EBS.
- Rivedi e ottimizza: esamina regolarmente il carico di lavoro per identificare e rimuovere eventuali risorse inutilizzate.

Risorse

Documenti correlati:

- [Ottimizzazione dell'infrastruttura AWS per la sostenibilità, Parte II: Archiviazione](#)
- [Come posso terminare risorse attive che non mi servono più sul mio Account AWS?](#)

Video correlati:

- [AWS re:Invent 2023 - Sustainable architecture: Past, present, and future](#)

- [AWS re:Invent 2022 - Preserving and maximizing the value of digital media assets using Amazon S3](#)
- [AWS re:Invent 2023 - Optimize costs in your multi-account environments](#)

SUS02-BP04 Ottimizzazione del posizionamento geografico dei carichi di lavoro in base ai requisiti di rete

Seleziona le sedi cloud e i servizi per il carico di lavoro per ridurre la distanza che il traffico di rete deve percorrere e diminuire così le risorse totali di rete richieste per supportare il carico di lavoro.

Anti-pattern comuni:

- Selezione della regione del carico di lavoro in base alla propria collocazione.
- Consolidamento di tutte le risorse del carico di lavoro in un'unica posizione geografica.
- Tutto il traffico passa attraverso i data center esistenti.

Vantaggi dell'adozione di questa best practice: il posizionamento di un carico di lavoro in prossimità dei relativi utenti garantisce la latenza più bassa e la contemporanea riduzione del trasferimento dei dati nella rete e dell'impatto ambientale.

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

Guida all'implementazione

L'infrastruttura Cloud AWS viene definita con opzioni diverse relative alle sedi, come aree geografiche, zone di disponibilità, gruppi di posizionamento e posizioni edge come [AWS Outposts](#) e [zone locali AWS](#). Queste opzioni relative alle sedi sono responsabili della gestione della connettività tra i componenti delle applicazioni, i servizi cloud, le reti edge e i data center on-premise.

Analizza i modelli di accesso alla rete nel tuo carico di lavoro per stabilire come usare queste opzioni relative alle sedi cloud e ridurre la distanza che il traffico di rete deve percorrere.

Passaggi dell'implementazione

- Analizza i modelli di accesso alla rete nel tuo carico di lavoro per capire come gli utenti usano la tua applicazione.
 - Usa strumenti di monitoraggio, come [Amazon CloudWatch](#) e [AWS CloudTrail](#), per raccogliere i dati sull'attività della rete.

- Analizza i dati per identificare il modello di accesso alla rete.
- Seleziona le regioni appropriate per l'implementazione del carico di lavoro in base ai seguenti elementi chiave:
 - Il tuo obiettivo di sostenibilità: spiegato nella [Selezione dell'area geografica](#).
 - Dove si trovano i tuoi dati: per le applicazioni a uso intensivo di dati, ad esempio applicazioni di big data e machine learning, il codice dell'applicazione deve essere eseguito il più vicino possibile ai dati.
 - Dove si trovano i tuoi utenti: per le applicazioni rivolte agli utenti, scegli un'area geografica (o più di una) vicina agli utenti del tuo carico di lavoro.
 - Altre limitazioni: considera le limitazioni relative a costi e conformità, come spiegato in [Cosa considerare quando si seleziona un'area geografica per i carichi di lavoro](#).
- Usa la cache locale o le [Soluzioni per la cache di AWS](#) per le risorse di frequente utilizzo per migliorare le performance, ridurre lo spostamento dei dati e minimizzare l'impatto ambientale.

Service	When to use
Amazon CloudFront	Usa per memorizzare nella cache contenuti statici come immagini, script e video, nonché contenuti dinamici come risposte API o applicazioni Web.
Amazon ElastiCache	Usa per memorizzare nella cache i contenuti per le applicazioni Web.
DynamoDB Accelerator	Usa per aggiungere accelerazione in memoria alle tabelle DynamoDB.

- Utilizza servizi in grado di supportarti nell'esecuzione del codice in posizioni più vicine agli utenti del carico di lavoro:

Service	When to use
Lambda@Edge	Usa per operazioni a uso intensivo di risorse di calcolo eseguite quando gli oggetti non si trovano nella cache.

Service	When to use
Funzioni Amazon CloudFront	Usa per casi d'uso semplici, ad esempio manipolazioni di risposte o richieste HTTP(s) che possono essere avviate da funzioni di breve durata.
AWS IoT Greengrass	Usa per eseguire la memorizzazione nella cache di risorse di calcolo, messaggistica e dati per i dispositivi connessi.

- Utilizza il pooling delle connessioni per consentire il loro riutilizzo e ridurre le risorse richieste.
- Utilizza archivi di dati distribuiti che non si affidano a connessioni persistenti e aggiornamenti sincroni per garantire coerenza e servire le popolazioni regionali.
- Sostituisci la capacità di rete statica preassegnata con una capacità dinamica condivisa e condividi l'impatto in termini di sostenibilità della capacità di rete con altri sottoscrittori.

Risorse

Documenti correlati:

- [Ottimizzazione dell'infrastruttura AWS per la sostenibilità, parte III: reti](#)
- [Documentazione Amazon ElastiCache](#)
- [Che cos'è Amazon CloudFront?](#)
- [Caratteristiche principali di Amazon CloudFront](#)
- [Infrastruttura globale AWS](#)
- [Zone locali AWS e AWS Outposts, scelta della giusta tecnologia per un carico di lavoro edge](#)
- [Gruppi di collocazione](#)
- [Zone locali AWS](#)
- [AWS Outposts](#)

Video correlati:

- [Demistificazione del trasferimento dei dati su AWS](#)
- [Scaling network performance on next-gen Amazon EC2 instances](#)

- [Video di presentazione delle zone locali AWS](#)
- [AWS Outposts: Overview and How it Works](#)
- [AWS re:Invent 2023 - A migration strategy for edge and on-premises workloads](#)
- [AWS re:Invent 2021: AWS Outposts: Spostamento dell'esperienza AWS in un ambiente on-premise](#)
- [AWS re:Invent 2020 - AWS Wavelength: Run apps with ultra-low latency at 5G edge](#)
- [AWS re:Invent 2022: Zone locali AWS: creazione di applicazioni per una posizione edge distribuita](#)
- [AWS re:Invent 2021: Creazione di siti Web a bassa latenza con Amazon CloudFront](#)
- [AWS re:Invent 2022: Miglioramento delle prestazioni e della disponibilità con AWS Global Accelerator](#)
- [AWS re:Invent 2022: Creazione di una rete WAN usando AWS](#)
- [AWS re:Invent 2020: Gestione del traffico globale con Amazon Route 53](#)

Esempi correlati:

- [Workshop di rete AWS](#)
- [Progettazione di architetture per la sostenibilità: riduzione al minimo dello spostamento dei dati tra reti](#)

SUS02-BP05 Ottimizzazione delle risorse dei membri del team in base alle attività eseguite

Ottimizza le risorse fornite ai membri del team per ridurre al minimo l'impatto sulla sostenibilità ambientale e supportare al tempo stesso le loro esigenze.

Anti-pattern comuni:

- Ignori l'impatto dei dispositivi utilizzati dai membri del tuo team sull'efficienza complessiva della tua applicazione cloud.
- Gestisci e aggiorni manualmente le risorse utilizzate dai membri del tuo team.

Vantaggi dell'adozione di questa best practice: se si ottimizzano le risorse dei membri del team si migliora l'efficienza generale delle applicazioni abilitate al cloud.

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

Guida all'implementazione

Identifica le risorse che i membri del tuo team usano per accedere ai tuoi servizi, il loro ciclo di vita atteso e l'impatto finanziario e di sostenibilità. Implementa strategie per ottimizzare queste risorse. Eseguì ad esempio operazioni complesse, come rendering e compilazione, su infrastrutture scalabili altamente utilizzate, invece che su sistemi per utenti singoli, sottoutilizzati e con un alto dispendio energetico.

Passaggi dell'implementazione

- Usa workstation efficienti dal punto di vista energetico: fornisci ai membri del team workstation e periferiche efficienti dal punto di vista energetico. Utilizza in questi dispositivi funzionalità di gestione dell'alimentazione efficienti, come la modalità di risparmio energetico, per ridurre il consumo di energia.
- Usa la virtualizzazione: utilizza i desktop virtuali e lo streaming delle applicazioni per limitare gli aggiornamenti e i requisiti dei dispositivi.
- Favorisci la collaborazione remota: incoraggia i membri del team a utilizzare strumenti di collaborazione remota come [Amazon Chime](#) o [AWS Wickr](#) per ridurre la necessità di spostamenti aziendali e le emissioni di carbonio associate.
- Usa software efficienti dal punto di vista energetico: fornisci ai membri del team software efficienti dal punto di vista energetico rimuovendo o disattivando funzionalità e processi non necessari.
- Gestisci i cicli di vita: valuta l'impatto di processi e sistemi sul ciclo di vita dei tuoi dispositivi e seleziona le soluzioni che riducono al minimo la necessità di sostituzione dei dispositivi, pur continuando a soddisfare i requisiti di business. Effettua regolarmente la manutenzione e l'aggiornamento delle workstation o del software per conservare e migliorare l'efficienza.
- Usa la gestione remota dei dispositivi: implementa la gestione remota dei dispositivi per ridurre gli spostamenti aziendali.
 - Gestione dei gruppi di nodi AWS Systems Manager è un'esperienza di interfaccia utente unificata che ti aiuta a gestire da remoto i nodi in esecuzione su AWS oppure on-premises.

Risorse

Documenti correlati:

- [Che cos'è Amazon WorkSpaces?](#)
- [Cost Optimizer per Amazon WorkSpaces](#)

- [Documentazione su Amazon AppStream 2.0](#)
- [NICE DCV](#)

Video correlati:

- [Gestire i costi per Amazon WorkSpaces su AWS](#)

SUS02-BP06 Implementazione del buffering o della limitazione (della larghezza di banda della rete) per ridurre la curva della domanda

Il buffering e la limitazione (della larghezza di banda della rete) riducono la curva delle richieste e la capacità fornita tramite provisioning per il tuo carico di lavoro.

Anti-pattern comuni:

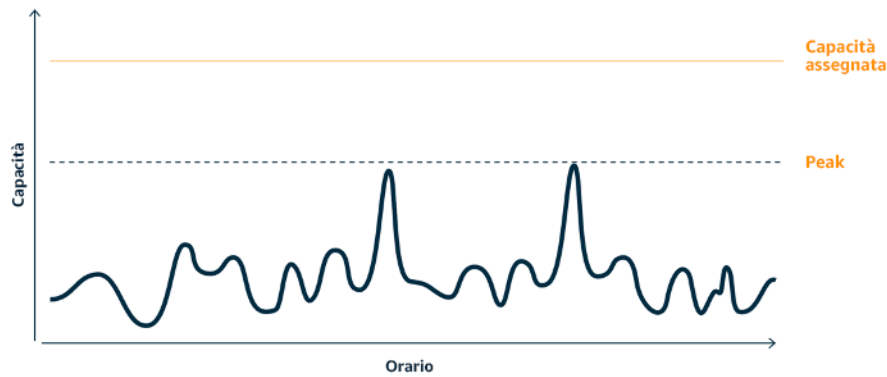
- Elabori immediatamente le richieste del client, anche se non è necessario.
- Non analizzi i requisiti relativi alle richieste dei clienti.

Vantaggi dell'azione di questa best practice: ridurre la curva della domanda per diminuire la capacità richiesta fornita tramite provisioning per il carico di lavoro. Ridurre la capacità fornita tramite provisioning significa ridurre il consumo di energia e contenere l'impatto ambientale.

Livello di rischio associato alla mancata adozione di questa best practice: basso

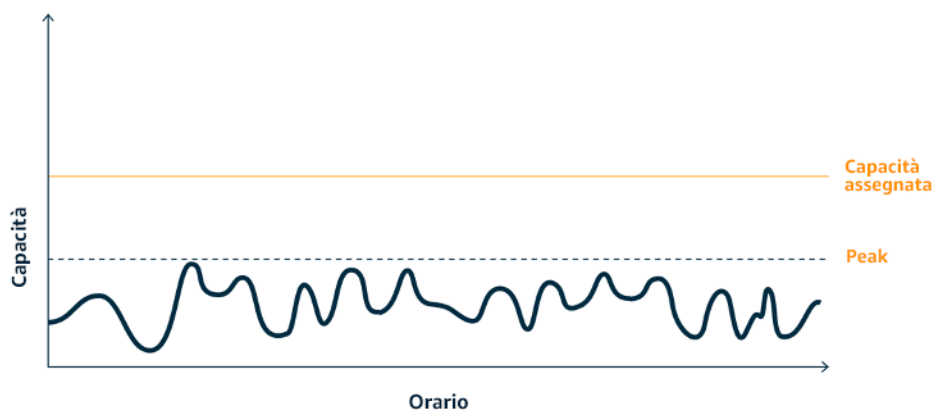
Guida all'implementazione

Diminuire la curva della domanda del carico di lavoro può aiutarti a ridurre la capacità fornita tramite provisioning di un carico di lavoro, oltre al suo impatto sull'ambiente. Supponiamo che un carico di lavoro abbia la curva della domanda mostrata nella figura qui sotto. Questo carico di lavoro presenta due picchi e per gestire tali picchi viene eseguito il provisioning della capacità di risorse mostrata dalla linea arancione. Le risorse e l'energia utilizzate per questo carico di lavoro non sono indicate nell'area sotto la curva della domanda, ma nell'area sotto la linea della capacità fornita, poiché per gestire questi due picchi è necessario eseguire il provisioning di tale capacità.



Curva della domanda con due picchi distinti che richiedono una capacità elevata.

Puoi usare il buffering o la limitazione (della larghezza di banda della rete) per modificare la curva della domanda e appianare i picchi, con conseguente diminuzione della capacità fornita tramite provisioning e consumo inferiore di energia. Implementa la limitazione (della larghezza di banda della rete) quando i client eseguono nuovi tentativi. Implementa il buffering per archiviare la richiesta e rinviare l'elaborazione a un secondo momento.



L'effetto della limitazione (della larghezza di banda della rete) sulla curva della domanda e la capacità fornita tramite provisioning.

Passaggi dell'implementazione

- Analizza le richieste del client per stabilire come rispondere. Le domande da considerare includono:
 - Questa richiesta può essere elaborata in modo asincrono?

- Il client ha la possibilità di ripetere i tentativi?
- Se il client ha la possibilità di ripetere i tentativi puoi implementare la limitazione (della larghezza di banda della rete), che indica alla sorgente che, se non è in grado di soddisfare la richiesta all'ora corrente, dovrebbe riprovare più tardi.
- Puoi usare [Amazon API Gateway](#) per implementare la limitazione (della larghezza di banda della rete).
- Per i client che non possono eseguire altri tentativi, è necessario implementare un buffer per ridurre i picchi della curva della domanda. Il buffering rinvia l'elaborazione delle richieste, consentendo alle applicazioni eseguite a velocità diverse di comunicare in modo efficace. Un approccio basato sul buffering impiega una coda o un flusso per l'accettazione dei messaggi dai produttori. I messaggi vengono letti ed elaborati dai consumatori e ciò consente ai messaggi di essere eseguiti alla velocità che soddisfa i requisiti aziendali del consumatore stesso.
- [Amazon Simple Queue Service\(Amazon SQS\)](#) è un servizio gestito che offre code che consentono a un singolo consumatore di leggere singoli messaggi.
- [Amazon Kinesis](#) offre un flusso che consente a più consumatori di leggere gli stessi messaggi.
- Analizza la domanda complessiva, la velocità di modifica e il tempo di risposta richiesto per determinare le dimensioni del throttling o del buffer richiesto.

Risorse

Documenti correlati:

- [Nozioni di base su Amazon SQS](#)
- [Integrazione dell'applicazione con code e messaggi](#)
- [Gestione e monitoraggio della limitazione delle API nei carichi di lavoro](#)
- [Throttling a tiered, multi-tenant REST API at scale using API Gateway](#)
- [Integrazione dell'applicazione con code e messaggi](#)

Video correlati:

- [AWS re:Invent 2022 - Application integration patterns for microservices](#)
- [AWS re:Invent 2023 - Smart savings: Amazon EC2 cost-optimization strategies](#)
- [AWS re:Invent 2023 - Advanced integration patterns & trade-offs for loosely coupled systems](#)

Software e architettura

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

Best practice

- [SUS03-BP01 Ottimizzazione del software e delle architetture per processi asincroni e pianificati](#)
- [SUS03-BP02 Rimozione o rifattorizzazione dei componenti dei carichi di lavoro con un utilizzo ridotto o assente](#)
- [Ottimizzazione delle aree di codice che consumano la maggior parte del tempo o delle risorse](#)
- [SUS03-BP04 Ottimizzazione dell'impatto su dispositivi e apparecchiature](#)
- [SUS03-BP05 Uso dei modelli e le architetture software che meglio supportano l'accesso ai dati e i modelli di archiviazione](#)

SUS03-BP01 Ottimizzazione del software e delle architetture per processi asincroni e pianificati

Utilizza modelli efficienti di software e di architettura, come quelli basati sulle code, per mantenere un utilizzo elevato e costante delle risorse distribuite.

Anti-pattern comuni:

- Provisioning di risorse in eccedenza per il carico di lavoro in cloud con lo scopo di far fronte a picchi di domanda imprevisti.
- Architettura non in grado di disaccoppiare i mittenti e i ricevitori di messaggi asincroni mediante un componente di messaggistica.

Vantaggi dell'adozione di questa best practice:

- Modelli efficienti di software e architettura riducono al minimo le risorse inutilizzate nel carico di lavoro e migliorano l'efficienza complessiva.
- È possibile dimensionare le risorse dedicate all'elaborazione indipendentemente dalla ricezione di messaggi asincroni.
- Grazie a un componente di messaggistica, i requisiti di disponibilità si attenuano e possono essere soddisfatti con un numero inferiore di risorse.

Livello di rischio associato alla mancata adozione di questa best practice: medio

Guida all'implementazione

Utilizza modelli di architettura efficienti, come l'[architettura basata su eventi](#), che consentono un utilizzo uniforme dei componenti e riducono al minimo il provisioning in eccedenza nel carico di lavoro. L'utilizzo di modelli architetturali efficienti riduce al minimo le risorse inattive a causa del mancato utilizzo dovuto alle variazioni della domanda nel tempo.

Comprendi i requisiti dei componenti del carico di lavoro e adotta modelli di architettura che aumentino l'utilizzo complessivo delle risorse. Ritira i componenti che non sono più necessari.

Passaggi dell'implementazione

- Analizza le esigenze del tuo carico di lavoro per determinare come rispondere a tali richieste.
- Per le richieste o i processi che non necessitano di risposte sincrone, utilizza architetture basate su code e worker a scalabilità automatica per massimizzare l'utilizzo. Ecco alcuni esempi in cui potresti prendere in considerazione un'architettura basata sulle code:

Queuing mechanism	Description
code di processi AWS Batch	I processi AWS Batch vengono inviati a una coda di processi, dove risiedono fino a quando non è possibile pianificare la loro esecuzione in un ambiente di calcolo.
Amazon Simple Queue Service e istanze spot Amazon EC2	Abbina Amazon SQS e istanze spot per realizzare un'architettura efficiente e in grado di tollerare i malfunzionamenti.

- Per le richieste o i processi che possono essere elaborati in qualsiasi momento, ottieni una maggiore efficienza utilizzando i meccanismi di pianificazione dell'elaborazione delle attività in blocco. Ecco alcuni esempi di meccanismi di pianificazione su AWS:

Scheduling mechanism	Description
Sistema di pianificazione Amazon EventBridge	Una funzionalità di Amazon EventBridge che consente di creare, eseguire e gestire attività pianificate su larga scala.
Pianificazione basata sul tempo di AWS Glue	Definisci una pianificazione in base al tempo per crawler e processi in AWS Glue.
Attività pianificate di Amazon Elastic Container Service (Amazon ECS)	Amazon ECS supporta la creazione di attività pianificate. Le attività pianificate utilizzano le regole di Amazon EventBridge per l'esecuzione in base a una pianificazione in risposta a un evento EventBridge.
Instance Scheduler	Configura le pianificazioni di avvio e arresto delle istanze Amazon EC2 e Amazon Relational Database Service.

- Se nella tua architettura utilizzi meccanismi di polling e webhook, sostituiscili con eventi. Utilizza [architetture basate su eventi](#) per realizzare carichi di lavoro efficienti.
- Approfitta dei servizi [serverless su AWS](#) per eliminare la necessità di provisioning in eccedenza sull'infrastruttura.
- Dimensiona in modo appropriato i singoli componenti dell'architettura per evitare la presenza di risorse inattive in attesa di input.
 - Puoi utilizzare i [suggerimenti per il dimensionamento appropriato in AWS Cost Explorer](#) o [AWS Compute Optimizer](#) per identificare le corrette opportunità di dimensionamento.
 - Per maggiori dettagli, consulta [Right Sizing: Provisioning Instances to Match Workloads](#).

Risorse

Documenti correlati:

- [Che cos'è Amazon Simple Queue Service?](#)
- [Che cos'è Amazon MQ?](#)
- [Dimensionamento basato su Amazon SQS](#)
- [Che cos'è AWS Step Functions?](#)
- [Che cos'è AWS Lambda?](#)
- [Utilizzo di AWS Lambda con Amazon SQS](#)
- [Che cos'è Amazon EventBridge?](#)
- [Managing Asynchronous Workflows with a REST API](#)

Video correlati:

- [AWS re:Invent 2023 - Navigating the journey to serverless event-driven architecture](#)
- [AWS re:Invent 2023 - Using serverless for event-driven architecture & domain-driven design](#)
- [AWS re:Invent 2023 - Advanced event-driven patterns with Amazon EventBridge](#)
- [AWS re:Invent 2023 - Sustainable architecture: Past, present, and future](#)
- [Asynchronous Message Patterns | AWS Events](#)

Esempi correlati:

- [Event-driven architecture with AWS Graviton Processors and Amazon EC2 Spot Instances](#)

SUS03-BP02 Rimozione o rifattorizzazione dei componenti dei carichi di lavoro con un utilizzo ridotto o assente

Elimina i componenti non utilizzati e non più necessari e rifattorizza quelli con scarso utilizzo per limitare lo spreco di risorse nel tuo carico di lavoro.

Anti-pattern comuni:

- Non verifichi con regolarità il livello di utilizzo dei singoli componenti del tuo carico di lavoro.
- Non verifichi e analizzi i consigli ricevuti dagli strumenti di dimensionamento AWS, ad esempio [AWS Compute Optimizer](#).

Vantaggi dell'adozione di questa best practice: se si eliminano i componenti non utilizzati si riducono gli sprechi e si migliora l'efficienza generale del carico di lavoro cloud.

Livello di rischio associato alla mancata adozione di questa best practice: medio

Guida all'implementazione

Esamina il tuo carico di lavoro per identificare componenti inattivi o non utilizzati. Si tratta di un processo di miglioramento iterativo che può essere attivato da cambiamenti della domanda o dal rilascio di un nuovo servizio cloud. Ad esempio, una riduzione significativa del runtime delle funzioni di [AWS Lambda](#) può essere un indicatore della necessità di diminuire la dimensione della memoria. Inoltre, quando AWS rilascia nuovi servizi e funzionalità, è possibile che i servizi ottimali e l'architettura per il carico di lavoro cambino.

Monitora continuamente l'attività del carico di lavoro e cerca le opportunità per migliorare il livello di utilizzo dei singoli componenti. Eliminando i componenti inattivi ed eseguendo attività di ridimensionamento, soddisfi i requisiti aziendali con il numero minimo di risorse cloud.

Passaggi dell'implementazione

- Redigi l'inventario delle tue risorse AWS. In AWS, puoi attivare [Esploratore di risorse AWS](#) per esplorare e organizzare le tue risorse AWS. Per maggiori dettagli, consulta [AWS re:Invent 2022 - How to manage resources and applications at scale on AWS](#).
- Monitora e acquisisci metriche di utilizzo per componenti strategici del tuo carico di lavoro (like l'utilizzo della CPU, l'utilizzo della memoria o la velocità di trasmissione effettiva nelle metriche [Amazon CloudWatch](#)).
- Individua i componenti inutilizzati o sottoutilizzati nell'architettura.
 - Per carichi di lavoro stabili, verifica gli strumenti di ridimensionamento AWS come [AWS Compute Optimizer](#) a intervalli regolari per individuare componenti inattivi, inutilizzati o sottoutilizzati.
 - Per carichi di lavoro effimeri, valuta metriche di utilizzo per identificare componenti inattivi, inutilizzati o sottoutilizzati.
- Ritira componenti e risorse associate (come le immagini Amazon ECR) che non sono più necessarie.
 - [Pulizia automatica delle immagini inutilizzate in Amazon ECR](#)
 - [Elimina volumi Amazon Elastic Block Store \(Amazon EBS\) utilizzando AWS Config e AWS Systems Manager](#)

- Rifattorizza o consolida i componenti sottoutilizzati con altre risorse per promuovere un utilizzo efficiente. Ad esempio, puoi eseguire il provisioning di più database di piccole dimensioni su una singola istanza di database [Amazon RDS](#) invece di eseguire database su singole istanze sottoutilizzate.
- Scopri le [risorse fornite dal tuo carico di lavoro per completare un'unità di lavoro](#).

Risorse

Documenti correlati:

- [AWS Trusted Advisor](#)
- [Che cos'è Amazon CloudWatch?](#)
- [Ridimensionamento corretto: provisioning delle istanze per soddisfare i carichi di lavoro](#)
- [Optimizing your cost with Rightsizing Recommendations](#)

Video correlati:

- [AWS re:Invent 2023 - Capacity, availability, cost efficiency: Pick three](#)

Esempi correlati:

- [Optimize Hardware Patterns and Observe Sustainability KPIs](#)

Ottimizzazione delle aree di codice che consumano la maggior parte del tempo o delle risorse

Ottimizza il codice eseguito all'interno di diversi componenti della tua architettura per ridurre l'utilizzo delle risorse e massimizzare al tempo stesso le prestazioni.

Anti-pattern comuni:

- Ignori l'ottimizzazione del codice per l'utilizzo delle risorse.
- In genere, rispondi ai problemi di performance aumentando le risorse.
- La revisione del codice e il processo di sviluppo non monitorano le modifiche a livello di performance.

Vantaggi dell'adozione di questa best practice: l'uso di un codice efficiente riduce la quantità di risorse utilizzate e migliora le performance.

Livello di rischio associato alla mancata adozione di questa best practice: medio

Guida all'implementazione

È fondamentale esaminare ogni area funzionale, incluso il codice per un'applicazione ideata nel cloud, per ottimizzare l'uso delle risorse e le performance. Monitora costantemente le performance del tuo carico di lavoro negli ambienti di sviluppo e produzione e identifica le opportunità per migliorare gli snippet di codice che comportano un utilizzo particolarmente elevato delle risorse. Adotta un processo di revisione con cadenza regolare per identificare i bug o gli anti-pattern all'interno del codice che utilizzano le risorse in modo non efficiente. Sfrutta algoritmi semplici ed efficienti che hanno gli stessi risultati per il tuo caso d'uso.

Passaggi dell'implementazione

- Usa un linguaggio di programmazione efficiente: utilizza un sistema operativo e un linguaggio di programmazione efficienti per il carico di lavoro. Per dettagli sui linguaggi di programmazione efficienti dal punto di vista delle risorse (incluso Rust), consulta [Sostenibilità con Rust](#).
- Usa un assistente per la scrittura del codice basato sull'IA: prendi in considerazione l'utilizzo di un assistente per la scrittura del codice basato sull'IA, ad esempio [Amazon CodeWhisperer](#), per scrivere il codice in modo efficiente.
- Automatizza le revisioni del codice: mentre sviluppi i tuoi carichi di lavoro, adotta un processo di revisione del codice automatizzato, per migliorare la qualità e identificare bug e anti-pattern.
 - [Automazione delle revisioni del codice con il Amazon CodeGuru Reviewer](#)
 - [Rilevare i bug concomitanti con Amazon CodeGuru](#)
 - [Aumentare la qualità del codice per le applicazioni Python con Amazon CodeGuru](#)
- Usa un profiler di codice: utilizza un profiler di codice per identificare le aree che impiegano più tempo o risorse e trasformale in obiettivi di ottimizzazione.
 - [Ridurre l'impatto ambientale della tua organizzazione con Amazon CodeGuru Profiler](#)
 - [Capire l'utilizzo della memoria nella tua applicazione Java con Amazon CodeGuru Profiler](#)
 - [Migliorare l'esperienza del cliente e ridurre i costi con Amazon CodeGuru Profiler](#)
- Monitora e ottimizza: utilizza le risorse di monitoraggio continuo per identificare i componenti con requisiti di risorse elevati o una configurazione non ottimale.

- Sostituisci gli algoritmi a uso intensivo di elaborazioni con una versione più semplice ed efficiente che produce gli stessi risultati.
- Rimuovi il codice non necessario, come quello relativo all'ordinamento e alla formattazione.
- Usa la rifattorizzazione o la trasformazione del codice: scopri la capacità di [trasformazione del codice di Amazon Q](#) per eseguire la manutenzione e gli aggiornamenti delle applicazioni.
- [Aggiorna le versioni linguistiche con la trasformazione del codice di Amazon Q](#)
- [AWS re:Invent 2023 - Automate app upgrades & maintenance using Amazon Q Code Transformation](#)

Risorse

Documenti correlati:

- [Che cos'è Amazon CodeGuru Profiler?](#)
- [Istanze FPGA](#)
- [SDK AWS su Strumenti per creare su AWS](#)

Video correlati:

- [Migliora l'efficienza del codice con Amazon CodeGuru Profiler](#)
- [AWS re:Invent 2023 - Best practices for Amazon CodeWhisperer](#)
- [Automatizza le revisioni del codice e i consigli sulle prestazioni dell'applicazione con Amazon CodeGuru](#)

Esempi correlati:

- [Optimizing Code with Amazon CodeGuru](#)

SUS03-BP04 Ottimizzazione dell'impatto su dispositivi e apparecchiature

Conoscere i dispositivi e le apparecchiature utilizzate nell'architettura e applicare strategie per ridurre il loro uso. Questo può ridurre l'impatto ambientale complessivo del tuo carico di lavoro cloud.

Anti-pattern comuni:

- Ignori l'impatto ambientale dei dispositivi utilizzati dai clienti.

- Gestisci e aggiorni manualmente le risorse utilizzate dai clienti.

Vantaggi dell'adozione di questa best practice: implementare modelli e funzionalità software ottimizzati per i dispositivi dei clienti può ridurre l'impatto ambientale complessivo del carico di lavoro del cloud.

Livello di rischio associato alla mancata adozione di questa best practice: medio

Guida all'implementazione

Implementare modelli e funzionalità software ottimizzati per i dispositivi dei clienti può ridurre l'impatto ambientale in diversi modi:

- Implementare nuove funzionalità compatibili con le versioni precedenti può ridurre il numero di sostituzioni hardware.
- Ottimizzare un'applicazione per un'esecuzione efficiente sui dispositivi può contribuire a ridurre l'utilizzo di energia ed estendere la durata della loro batteria (se sono alimentati tramite batteria).
- Ottimizzare un'applicazione per i dispositivi significa anche ridurre il trasferimento dei dati sulla rete.

Conoscere i dispositivi e l'attrezzatura utilizzati nella tua architettura, il loro ciclo di vita atteso e l'impatto della sostituzione di tali componenti. Implementare modelli e funzionalità software che possono contribuire a ridurre l'uso di energia da parte del dispositivo, la necessità da parte dei clienti di sostituirlo e anche di eseguire l'aggiornamento manuale.

Passaggi dell'implementazione

- Redigi l'inventario: esegui l'inventario dei dispositivi utilizzati nell'architettura. I dispositivi possono essere cellulari, tablet, dispositivi IOT, illuminazione smart o persino dispositivi smart in una fabbrica.
- Usa dispositivi efficienti dal punto di vista energetico: prendi in considerazione l'utilizzo di dispositivi ad alta efficienza energetica nell'architettura. Utilizza le configurazioni di gestione dell'alimentazione sui dispositivi per accedere alla modalità di risparmio energetico quando non sono in uso.
- Esegui applicazioni efficienti: ottimizza l'applicazione in esecuzione sui dispositivi:
 - Usa strategie come l'esecuzione di attività in background per ridurre l'uso di energia.

- Prendi in considerazione la larghezza di banda e la latenza della rete durante la creazione di payload e implementa funzionalità che consentano alle tue applicazioni di lavorare bene anche in presenza di una larghezza di banda ridotta e di link ad alta latenza.
- Converti payload e file in formati ottimizzati richiesti dai dispositivi. Ad esempio, puoi usare [Amazon Elastic Transcoder](#) o [AWS Elemental MediaConvert](#) per convertire file di media digitali di grandi dimensioni e di qualità elevata in formati che gli utenti possono riprodurre su dispositivi mobili, tablet, browser web e televisioni connesse.
- Esegui attività a elevata intensità computazionale lato server (come, ad esempio, il rendering delle immagini) oppure usa lo streaming delle applicazioni per migliorare l'esperienza utente sui dispositivi di versioni precedenti.
- Esegui la segmentazione e la paginazione dell'output, soprattutto per le sessioni interattive, per gestire i payload e limitare i requisiti di archiviazione in locale.
- Coinvolgi i fornitori: collabora con i fornitori di dispositivi che utilizzano materiali sostenibili e forniscono trasparenza nelle catene di approvvigionamento e nelle certificazioni ambientali.
- Usa gli aggiornamenti via etere (OTA): utilizza il meccanismo automatico via etere (OTA) per implementare gli aggiornamenti su uno o più dispositivi.
 - Puoi usare una [pipeline CI/CD](#) per aggiornare le applicazioni mobili.
 - Puoi usare [AWS IoT Device Management](#) per gestire da remoto dispositivi connessi su scala.
- Usa device farm gestite: per testare nuove funzionalità e aggiornamenti, utilizza device farm gestite con set di hardware rappresentativi e itera lo sviluppo per ottimizzare i dispositivi supportati. Per ulteriori dettagli, consulta [SUS06-BP04 Utilizzo di device farm gestite per i test](#).
- Monitora e migliora in modo continuo: monitora il consumo energetico dei dispositivi per identificare le aree di miglioramento. Utilizza le nuove tecnologie o best practice per migliorare l'impatto ambientale di questi dispositivi.

Risorse

Documenti correlati:

- [What is AWS Device Farm?](#)
- [AppStream 2.0 Documentation](#)
- [NICE DCV](#)
- [Tutorial OTA per l'aggiornamento del firmware su dispositivi che eseguono FreeRTOS](#)
- [Optimizing Your IoT Devices for Environmental Sustainability](#)

Video correlati:

- [AWS re:Invent 2023 - Improve your mobile and web app quality using AWS Device Farm](#)

SUS03-BP05 Uso dei modelli e le architetture software che meglio supportano l'accesso ai dati e i modelli di archiviazione

Scopri come i dati vengono utilizzati all'interno del tuo carico di lavoro, consumati dagli utenti, trasferiti e archiviati. Usa architetture e modelli software in grado di supportare al meglio l'accesso ai dati e l'archiviazione per ridurre le risorse di elaborazione, rete e storage richieste dal carico di lavoro.

Anti-pattern comuni:

- Ritieni che tutti i carichi di lavoro abbiano modelli di accesso e archiviazione dei dati simili.
- Utilizzi un solo livello di archiviazione, presupponendo che tutti i carichi di lavoro rientrino in tale livello.
- Ritieni che gli schemi di accesso ai dati rimarranno coerenti nel tempo.
- La tua architettura supporta una potenziale espansione elevata dell'accesso ai dati, con conseguente inattività delle risorse per la maggior parte del tempo.

Vantaggi dell'adozione di questa best practice: selezionando e ottimizzando la tua architettura in base all'accesso ai dati e ai modelli di archiviazione diminuirà la complessità dello sviluppo e aumenterà l'utilizzo complessivo. Capire quando utilizzare le tabelle globali, il partizionamento dei dati e la memorizzazione nella cache, ti aiuterà a ridurre i costi operativi e a effettuare il dimensionamento in base alle esigenze del carico di lavoro.

Livello di rischio associato alla mancata adozione di questa best practice: medio

Guida all'implementazione

Usa modelli di software e architetture che siano quanto più in linea con le caratteristiche dei tuoi dati e i modelli di accesso. Ad esempio, usa [un'architettura di dati moderni su AWS](#) che ti consenta di utilizzare servizi dedicati ottimizzati per i tuoi casi d'uso di analisi specifici. Questi modelli di architettura consentono un'elaborazione efficiente dei dati e riducono l'utilizzo delle risorse.

Passaggi dell'implementazione

- Analizza le caratteristiche dei dati e i modelli di accesso per individuare la configurazione corretta per le tue risorse cloud. Gli aspetti chiave da considerare includono:
 - Tipi di dati: strutturati, semi-strutturati, non strutturati
 - Crescita dei dati: delimitati, non delimitati
 - Durabilità dei dati: persistenti, effimeri, transitori
 - Modelli di accesso: letture o scritture, frequenza di aggiornamento, con picchi o costante
- Usa tipi di architetture che meglio supportano l'accesso ai dati e i modelli di archiviazione.
 - [Pattern per consentire la persistenza dei dati](#)
 - [Progettiamo! Architetture dei dati moderne](#)
 - [Database su AWS: lo strumento più adatto per ciascun processo](#)
- Sfrutta le tecnologie che lavorano in modo nativo con i dati compressi.
 - [Formati file di supporto alla compressione di Athena](#)
 - [Opzioni di formato per input e output ETL in AWS Glue](#)
 - [Caricamento di file di dati compressi da Amazon S3 con Amazon Redshift](#)
- Usa [servizi di analisi](#) per l'elaborazione dei dati nella tua architettura. Per informazioni dettagliate sui servizi di analisi personalizzati di AWS, consulta [AWS re:Invent 2022 - Building modern data architectures on AWS](#).
- Utilizza il motore del database che meglio supporta il modello di query dominante. Gestisci gli indici di database per garantire un'esecuzione efficiente delle query. Per ulteriori dettagli, consulta [Database AWS](#) e [AWS re:Invent 2022 - Modernize apps with purpose-built databases](#).
- Seleziona protocolli di rete che riducano la quantità di capacità di rete utilizzata dalla tua architettura.

Risorse

Documenti correlati:

- [COPY dai formati dei dati in colonne con Amazon Redshift](#)
- [Convertire il formato dei record di input in Firehose](#)
- [Migliora le prestazioni delle query su Amazon Athena con una conversione ai formati in colonne](#)
- [Monitoraggio del carico del database con Performance Insights su Amazon Aurora](#)

- [Monitoraggio del carico del database con Performance Insights su Amazon RDS](#)
- [Classe di storage Amazon S3 Intelligent-Tiering](#)
- [Build a CQRS event store with Amazon DynamoDB](#)

Video correlati:

- [AWS re:Invent 2022 - Building data mesh architectures on AWS](#)
- [AWS re:Invent 2023 - Deep dive into Amazon Aurora and its innovations](#)
- [AWS re:Invent 2023 - Improve Amazon EBS efficiency and be more cost-efficient](#)
- [AWS re:Invent 2023 - Optimizing storage price and performance with Amazon S3](#)
- [AWS re:Invent 2023 - Building and optimizing a data lake on Amazon S3](#)
- [AWS re:Invent 2023 - Advanced event-driven patterns with Amazon EventBridge](#)

Esempi correlati:

- [AWS Purpose Built Databases Workshop](#)
- [AWS Modern Data Architecture Immersion Day](#)
- [Build a Data Mesh on AWS](#)

Gestione dati

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

Best practice

- [SUS04-BP01 Implementazione di una policy di classificazione dei dati](#)
- [SUS04-BP02 Utilizzo di tecnologie che supportano l'accesso ai dati e i modelli di archiviazione](#)
- [SUS04-BP03 Utilizzo delle policy per gestire il ciclo di vita dei set di dati](#)
- [SUS04-BP04 Utilizzo dell'elasticità e dell'automazione per espandere lo storage a blocchi o il file system](#)
- [SUS04-BP05 Eliminazione dei dati ridondanti o non necessari](#)

- [SUS04-BP06 Utilizzo di file system condivisi o archiviazione per accedere a dati comuni](#)
- [SUS04-BP07 Riduzione al minimo dello spostamento di dati tra reti](#)
- [SUS04-BP08 Backup dei dati solo quando sono difficili da ricreare](#)

SUS04-BP01 Implementazione di una policy di classificazione dei dati

Classifica i dati per capire le criticità rispetto ai risultati aziendali e scegli il livello di archiviazione ad alta efficienza corretto per le tue informazioni.

Anti-pattern comuni:

- Non identifichi asset di dati con caratteristiche simili (come sensibilità, criticità aziendale o requisiti normativi) che vengono elaborati o archiviati.
- Non hai implementato un catalogo di dati per eseguire l'inventario dei tuoi asset.

Vantaggi derivanti dall'adozione di questa best practice: l'implementazione di una policy di classificazione dei dati ti consente di stabilire il livello di archiviazione per i dati più efficiente dal punto di vista energetico.

Livello di rischio associato alla mancata adozione di questa best practice: medio

Guida all'implementazione

La classificazione dei dati comporta l'identificazione dei tipi di dati elaborati e archiviati in un sistema informativo di proprietà o gestito da un'organizzazione. Inoltre, è necessario stabilire la criticità dei dati e il probabile impatto di una compromissione, perdita o uso improprio dei dati.

Implementare la policy di classificazione dei dati partendo dall'uso contestuale dei dati e creando uno schema di categorizzazione che tenga conto del livello di criticità di un determinato set di dati per le operazioni dell'organizzazione.

Passaggi dell'implementazione

- Esegui l'inventario dei dati: redigi l'inventario dei vari tipi di dati esistenti per il carico di lavoro.
- Raggruppa i dati: determina la criticità, la riservatezza, l'integrità e la disponibilità dei dati in base al rischio per l'organizzazione. Utilizza questi requisiti per raggruppare i dati in uno dei livelli di classificazione dei dati adottati. Come esempio vedi [Quattro semplici passaggi per classificare i tuoi dati e proteggere la tua startup](#).

- Definisci i livelli di classificazione dei dati e le policy: per ogni gruppo di dati, definisci il livello di classificazione dei dati, ad esempio pubblico o riservato, e le policy di gestione. Applica ai dati i tag adeguati. Per maggiori dettagli sulle categorie di classificazione dei dati, consulta il whitepaper sulla classificazione dei dati.
- Rivedi periodicamente: esamina e controlla periodicamente il tuo ambiente per individuare i dati senza tag e non classificati. Usa l'automazione per identificare questi dati, classificandoli e applicando i tag in modo appropriato. Come esempio vedi [Catalogo dati e crawler in AWS Glue](#).
- Crea un catalogo dati: definisci un catalogo dati che fornisca funzionalità di audit e governance.
- Documenta: crea i documenti per comunicare le policy di classificazione dei dati e le procedure di gestione per ogni classe di dati.

Risorse

Documenti correlati:

- [Utilizzo di Cloud AWS per supportare la classificazione dei dati](#)
- [Policy di tag di AWS Organizations](#)

Video correlati:

- [AWS re:Invent 2022 - Enabling agility with data governance on AWS](#)
- [AWS re:Invent 2023 - Data protection and resilience with AWS storage](#)

SUS04-BP02 Utilizzo di tecnologie che supportano l'accesso ai dati e i modelli di archiviazione

Usa tecnologie di archiviazione in grado di supportare al meglio il modo in cui viene effettuato l'accesso ai dati e come vengono archiviati per ridurre la quantità di risorse assegnate e supportare al tempo stesso il tuo carico di lavoro.

Anti-pattern comuni:

- Ritieni che tutti i carichi di lavoro abbiano modelli di accesso e archiviazione dei dati simili.
- Utilizzi un solo livello di archiviazione, presupponendo che tutti i carichi di lavoro rientrino in tale livello.
- Ritieni che gli schemi di accesso ai dati rimarranno coerenti nel tempo.

Vantaggi derivanti da questa best practice: selezionare e ottimizzare le tecnologie di archiviazione in base all'accesso ai dati e ai modelli di archiviazione ti consentirà di ridurre le risorse cloud richieste per soddisfare le tue esigenze aziendali e migliorare l'efficienza generale del tuo carico di lavoro cloud.

Livello di rischio associato alla mancata adozione di questa best practice: basso

Guida all'implementazione

Seleziona la soluzione di archiviazione più adatta ai tuoi modelli di accesso. In alternativa, puoi modificarli affinché siano in linea con la soluzione di archiviazione, allo scopo di ottimizzare l'efficienza delle prestazioni.

Passaggi dell'implementazione

- Esamina le caratteristiche dei dati e dell'accesso: valuta le caratteristiche dei dati e il modello di accesso per raccogliere gli aspetti chiave delle tue esigenze di archiviazione. Gli aspetti chiave da considerare includono:
 - Tipi di dati: strutturati, semi-strutturati, non strutturati
 - Crescita dei dati: delimitati, non delimitati
 - Durabilità dei dati: persistenti, effimeri, transitori
 - Modelli di accesso: letture o scritture, frequenza, con picchi o costante
- Scegli la giusta tecnologia di archiviazione: migra i dati alla tecnologia di archiviazione appropriata che supporta le caratteristiche dei dati e il modello di accesso che usi. Ecco alcuni esempi di tecnologie di archiviazione AWS e delle loro caratteristiche chiave:

Type	Technology	Key characteristics
Archiviazione di oggetti	Amazon S3	Un servizio di archiviazione di oggetti con scalabilità illimitata, elevata disponibilità e più opzioni di accessibilità. Il trasferimento di oggetti e l'accesso a oggetti in e fuori da Amazon S3 possono utilizzare un servizio, ad esempio Transfer Accelerat

Type	Technology	Key characteristics
		ion o Access Points per supportare la tua posizione , le esigenze di sicurezza e i modelli di accesso.
Archiviazione	Amazon S3 Glacier	Classe di archiviazione di Amazon S3 creata per l'archiviazione dei dati.
File system condiviso	Amazon Elastic File System (Amazon EFS)	File system montabile a cui possono accedere più tipi di soluzioni di calcolo. Amazon EFS aumenta o riduce automaticamente lo spazio di archiviazione, mentre le relative prestazioni sono ottimizzate in modo da offrire costantemente latenze basse.
File system condiviso	Amazon FSx	Sviluppato sulle più recenti soluzioni di elaborazione AWS per supportare quattro file system comunemente usati: NetApp ONTAP, OpenZFS, Windows File Server e Lustre. Amazon FSx latenza, throughput e IOPS variano per file system e dovrebbero essere presi in considerazione quando selezioni il file system corretto per le esigenze del tuo carico di lavoro.

Type	Technology	Key characteristics
Archiviazione a blocchi	Amazon Elastic Block Store (Amazon EBS)	Servizio di archiviazione a blocchi scalabile e ad alte prestazioni progettato per Amazon Elastic Compute Cloud (Amazon EC2). Amazon EBS include l'archiviazione supportata da SSD per carichi di lavoro transazionali e a uso intensivo di IOPS, nonché archiviazione supportata da HDD per carichi di lavoro con uso intensivo della velocità di trasmissione effettiva.
Database relazionale	Amazon Aurora , Amazon RDS , Amazon Redshift	Progettati per supportare le transazioni ACID (atomicità, coerenza, isolamento, durabilità) e per mantenere l'integrità referenziale e una solida coerenza dei dati. Molte applicazioni tradizionali e sistemi Enterprise Resource Planning (ERP), Customer Relationship Management (CRM) ed e-commerce utilizzano database relazionali per archiviare i propri dati.

Type	Technology	Key characteristics
Database chiave-valore	Amazon DynamoDB	Ottimizzati per schemi di accesso di uso comune, in genere per archiviare e recuperare grandi volumi di dati. Le app Web dal traffico elevato, i sistemi di e-commerce e le applicazioni di videogiochi sono casi d'uso tipici dei database chiave-valore.

- Automatizza l'allocazione dello spazio di archiviazione: per i sistemi di archiviazione con dimensione fissa, come Amazon EBS o Amazon FSx, monitora lo spazio di archiviazione disponibile e automatizza l'allocazione dell'archiviazione al raggiungimento di una soglia. È possibile sfruttare Amazon CloudWatch al fine di raccogliere e analizzare metriche diverse per [Amazon EBS](#) e [Amazon FSx](#).
- Scegli la classe di archiviazione corretta: seleziona la classe di archiviazione appropriata per i tuoi dati.
 - Le classi di archiviazione Amazon S3 possono essere configurate a livello di oggetto. Un singolo bucket può contenere oggetti archiviati per tutte le classi di archiviazione.
 - Si possono utilizzare le policy del ciclo di vita Amazon S3 per passare automaticamente gli oggetti tra le classi di archiviazione oppure rimuovere i dati senza modifiche all'applicazione. In generale, devi raggiungere un equilibrio tra efficienza delle risorse, latenza di accesso e affidabilità, quando consideri questi meccanismi di storage.

Risorse

Documenti correlati:

- [Tipi di volume Amazon EBS](#)
- [Archivio dell'istanza Amazon EC2](#)
- [Intelligent Tiering di Amazon S3](#)
- [Caratteristiche di I/O Amazon EBS](#)
- [Utilizzo delle classi di archiviazione di Amazon S3](#)

- [Che cos'è Amazon S3 Glacier?](#)

Video correlati:

- [AWS re:Invent 2023 - Improve Amazon EBS efficiency and be more cost-efficient](#)
- [AWS re:Invent 2023 - Optimizing storage price and performance with Amazon S3](#)
- [AWS re:Invent 2023 - Building and optimizing a data lake on Amazon S3](#)
- [AWS re:Invent 2022 - Building modern data architectures on AWS](#)
- [AWS re:Invent 2022 - Modernize apps with purpose-built databases](#)
- [AWS re:Invent 2022 - Building data mesh architectures on AWS](#)
- [AWS re:Invent 2023 - Deep dive into Amazon Aurora and its innovations](#)
- [AWS re:Invent 2023 - Advanced data modeling with Amazon DynamoDB](#)

Esempi correlati:

- [Esempi di Amazon S3](#)
- [AWS Purpose Built Databases Workshop](#)
- [Databases for Developers](#)
- [AWS Modern Data Architecture Immersion Day](#)
- [Build a Data Mesh on AWS](#)

SUS04-BP03 Utilizzo delle policy per gestire il ciclo di vita dei set di dati

Gestisci il ciclo di vita di tutti i tuoi dati e applica in automatico le cancellazioni per ridurre i requisiti totali di archiviazione del tuo carico di lavoro.

Anti-pattern comuni:

- Cancellazione manuale dei dati.
- Conservazione di tutti i dati del carico di lavoro.
- Mancato spostamento dei dati su livelli di archiviazione più efficienti dal punto di vista energetico in base ai requisiti di conservazione e accesso.

Vantaggi dell'adozione di questa best practice: l'utilizzo delle policy per il ciclo di vita dei dati garantisce un accesso e una conservazione efficienti dei dati in un carico di lavoro.

Livello di rischio associato alla mancata adozione di questa best practice: medio

Guida all'implementazione

I set di dati presentano solitamente requisiti di conservazione e accesso che cambiano durante il loro ciclo di vita. Ad esempio, l'applicazione potrebbe avere bisogno di accedere frequentemente ad alcuni set di dati per un periodo di tempo limitato. In seguito, questi set di dati vengono consultati di rado.

Per gestire in modo efficiente i set di dati durante il loro ciclo di vita, è necessario configurare le policy per il ciclo di vita, ovvero le regole che definiscono la gestione dei set di dati.

Con le regole di configurazione del ciclo di vita, è possibile indicare al servizio di archiviazione di trasferire un set di dati a livelli di archiviazione più efficienti dal punto di vista energetico, di archivarlo o di eliminarlo.

Passaggi dell'implementazione

- [Classifica i set di dati del carico di lavoro.](#)
- Definisci le procedure di gestione per ogni classe di dati.
- Imposta policy automatizzate per il ciclo di vita per applicare le regole correlate. Ecco alcuni esempi di come impostare policy automatizzate per il ciclo di vita di diversi servizi di archiviazione di AWS:

Storage service	How to set automated lifecycle policies
Amazon S3	<p>È possibile utilizzare Amazon S3 Lifecycle per gestire gli oggetti durante il loro ciclo di vita. Se gli schemi di accesso sono sconosciuti, mutevoli o imprevedibili, è possibile utilizzare Amazon S3 Intelligent-Tiering, che monitora gli schemi di accesso e sposta automaticamente gli oggetti che non hanno fatto registrare accessi a livelli di accesso più economici. Puoi sfruttare i parametri di Amazon S3 Storage Lens per identificare le</p>

Storage service	How to set automated lifecycle policies
	opportunità di ottimizzazione e le lacune nella gestione del ciclo di vita.
Amazon Elastic Block Store	Puoi utilizzare Amazon Data Lifecycle Manager per automatizzare la creazione, la conservazione e l'eliminazione degli snapshot Amazon EBS e delle AMI basate su Amazon EBS.
Amazon Elastic File System	La gestione del ciclo di vita di Amazon EFS gestisce automaticamente l'archiviazione dei file per i tuoi file system.
Amazon Elastic Container Registry	Le policy per il ciclo di vita di Amazon ECR automatizzano la pulizia delle immagini dei container, facendo scadere le immagini in base all'età o al numero di copie.
AWS Elemental MediaStore	Si può usare una policy per il ciclo di vita di un oggetto che regola la durata di conservazione degli oggetti nel container MediaStore.

- Elimina i volumi inutilizzati, gli snapshot e i dati che hanno superato il periodo di conservazione. Sfrutta le funzionalità native del servizio, come il [Time To Live di Amazon DynamoDB](#) o la [conservazione dei log di Amazon CloudWatch](#) per programmare l'eliminazione.
- Aggrega e comprimi i dati quando possibile in base alle regole del ciclo di vita.

Risorse

Documenti correlati:

- [Ottimizza le regole del ciclo di vita di Amazon S3 con l'analisi delle classi di archiviazione di Amazon S3](#)
- [Valutazione delle risorse con Regole di AWS Config](#)

Video correlati:

- [AWS re:Invent 2021 - Amazon S3 Lifecycle best practices to optimize your storage spend](#)
- [AWS re:Invent 2023 - Optimizing storage price and performance with Amazon S3](#)
- [Semplifica il ciclo di vita dei dati e ottimizza i costi di archiviazione con Amazon S3 Lifecycle](#)
- [Riduci i costi di archiviazione con Amazon S3 Storage Lens](#)

SUS04-BP04 Utilizzo dell'elasticità e dell'automazione per espandere lo storage a blocchi o il file system

Usa l'elasticità e l'automazione per espandere lo storage a blocchi o il file system con l'aumento dei dati per ridurre l'archiviazione totale oggetto di provisioning.

Anti-pattern comuni:

- Acquisti uno storage a blocchi di grandi dimensioni o un file system per necessità future.
- Esegui un provisioning eccessivo delle operazioni di input e output al secondo (IOPS) del tuo file system.
- Non monitori l'utilizzo dei volumi di dati.

Vantaggi dell'adozione di questa best practice: ridurre il provisioning eccessivo per il sistema di archiviazione significa ridurre le risorse inattive e migliorare l'efficienza complessiva del tuo carico di lavoro.

Livello di rischio associato alla mancata adozione di questa best practice: medio

Guida all'implementazione

Crea storage a blocchi e file system con l'allocazione delle dimensioni, la velocità di trasmissione effettiva e la latenza adeguate al tuo carico di lavoro. Usa l'elasticità e l'automazione per espandere lo storage a blocchi o il file system con l'aumento dei dati per evitare un provisioning eccessivo per questi servizi di archiviazione.

Passaggi dell'implementazione

- Per i sistemi di storage di dimensioni fisse, come [Amazon EBS](#), assicurati di monitorare la quantità di archiviazione utilizzata rispetto alle dimensioni complessive dell'archiviazione e di creare, se possibile, un'automazione per aumentarne le dimensioni quando si raggiunge una soglia.

- Utilizza volumi elastici e servizi di dati a blocchi gestiti per automatizzare l'allocazione di archivi aggiuntivi man mano che i dati persistenti aumentano. Come esempio puoi usare i [Volumi elastici Amazon EBS](#) per modificare le dimensioni dei volumi, il tipo di volume o adeguare le performance dei tuo volumi Amazon EBS.
- Scegli la classe di archiviazione corretta, le performance e la velocità di trasmissione effettiva per il tuo file system per rispondere alle esigenze della tua azienda, senza eccedere.
 - [Performance Amazon EFS](#)
 - [Performance dei volumi Amazon EBS sulle istanze Linux](#)
- Imposta i livelli target di utilizzo per i volumi di dati e ridimensiona i volumi al di fuori degli intervalli previsti.
- Dimensiona i volumi di sola lettura per adattarli ai dati.
- Migra i dati su archivi oggetti per evitare il provisioning di capacità eccessive da dimensioni di volumi fisse su archiviazioni a blocchi.
- Esamina regolarmente i volumi elastici e i file system per terminare i volumi inattivi e ridurre i volumi con un provisioning eccessivo per adattarli alla dimensione corrente dei dati.

Risorse

Documenti correlati:

- [Extend the file system after resizing an EBS volume](#)
- [Modify a volume using Amazon EBS Elastic Volumes](#)
- [Documentazione Amazon FSx](#)
- [Che cos'è Amazon Elastic File System?](#)

Video correlati:

- [Approfondimento sui volumi elastici di Amazon EBS](#)
- [Strategie Amazon EBS e di ottimizzazione degli snapshot per performance migliori e risparmio sui costi](#)
- [Ottimizzare Amazon EFS per costi e performance, usando le best practice](#)

SUS04-BP05 Eliminazione dei dati ridondanti o non necessari

Elimina i dati non necessari o ridondanti per ridurre al minimo le risorse di archiviazione necessarie per memorizzare i set di dati.

Anti-pattern comuni:

- Duplicazione dei dati che possono essere facilmente recuperati o ricreati.
- Backup di tutti i dati senza prenderne in considerazione la criticità.
- Cancellazione dei dati eseguita in modo irregolare, in occasione di eventi operativi o non eseguita affatto.
- Archiviazione dei dati in modo ridondante, indipendentemente dall'affidabilità del servizio di archiviazione.
- Attivazione del controllo delle versioni di Amazon S3 senza alcuna giustificazione aziendale.

Vantaggi dell'adozione di questa best practice: la rimozione dei dati non necessari riduce le dimensioni dello spazio di archiviazione necessario per il carico di lavoro e il relativo impatto ambientale.

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

Guida all'implementazione

Non memorizzare i dati che non ti servono. Automatizza l'eliminazione dei dati non necessari. Utilizza tecnologie di backup che deduplicano i dati a livello di file e blocco. Sfrutta le funzionalità native di replica e ridondanza dei dati dei servizi.

Passaggi dell'implementazione

- Valuta se è possibile evitare la memorizzazione dei dati utilizzando set di dati esistenti disponibili pubblicamente in [AWS Data Exchange](#) e [Open Data su AWS](#).
- Utilizza meccanismi che possano deduplicare i dati a livello di blocco e oggetto. Ecco alcuni esempi di come deduplicare i dati su AWS:

Storage service	Deduplication mechanism
Amazon S3	Utilizza AWS Lake Formation FindMatches per individuare i record corrispondenti in un

Storage service	Deduplication mechanism
	set di dati (compresi quelli senza identificatori), utilizzando il nuovo FindMatches ML Transform.
Amazon FSx	Usa la deduplicazione dei dati su Amazon FSx per Windows.
Snapshot di Amazon Elastic Block Store	Gli snapshot sono backup incrementali, il che significa che vengono salvati solo i blocchi sul dispositivo che sono stati modificati dopo lo snapshot più recente.

- Analizza l'accesso ai dati per identificare quelli non necessari. Automatizza le policy per il ciclo di vita. Sfrutta le caratteristiche native del servizio, come il [Time To Live di Amazon DynamoDB](#), [Amazon S3 Lifecycle](#) o la [conservazione dei log di Amazon CloudWatch](#) per l'eliminazione.
- Utilizza le funzionalità di virtualizzazione dei dati di AWS per mantenere i dati sul loro sistema di origine ed evitare la loro duplicazione.
 - [Cloud Native Data Virtualization on AWS](#)
 - [Optimize Data Pattern Using Amazon Redshift Data Sharing](#)
- Utilizza una tecnologia di backup in grado di eseguire backup incrementali.
- Per raggiungere i tuoi obiettivi di persistenza, sfrutta l'affidabilità di [Amazon S3](#) e la [replica di Amazon EBS](#) invece di tecnologie da gestire in autonomia (come i dischi RAID).
- Centralizza i log e traccia i dati, deduplica le voci di log identiche e stabilisci meccanismi per ottimizzarne la verbosità quando necessario.
- Popola in anticipo le cache solo quando è necessario.
- Definisci il monitoraggio e l'automazione della cache per ridimensionarla in base alle esigenze.
- Rimuovi le implementazioni e le risorse obsolete dagli archivi di oggetti e dalle cache edge durante la distribuzione di nuove versioni del carico di lavoro.

Risorse

Documenti correlati:

- [Modifica la conservazione dei dati di log in CloudWatch Logs](#)

- [Deduplicazione dei dati su Amazon FSx per Windows File Server](#)
- [Funzionalità di Amazon FSx per ONTAP, compresa la deduplicazione dei dati](#)
- [Invalidazione dei file su Amazon CloudFront](#)
- [Utilizzo di AWS Backup per il backup e il ripristino dei file system di Amazon EFS](#)
- [Che cos'è Amazon CloudWatch Logs?](#)
- [Working with backups on Amazon RDS](#)
- [Integrate and deduplicate datasets using AWS Lake Formation](#)

Video correlati:

- [Amazon Redshift Data Sharing Use Cases](#)

Esempi correlati:

- [Come faccio ad analizzare i miei log di accesso al server Amazon S3 utilizzando Amazon Athena?](#)

SUS04-BP06 Utilizzo di file system condivisi o archiviazione per accedere a dati comuni

Adotta file system condivisi o l'archiviazione per evitare duplicazioni di dati e abilitare un'infrastruttura più efficiente per il tuo carico di lavoro.

Anti-pattern comuni:

- Esegui il provisioning dell'archiviazione per ogni singolo client.
- Non scolleghi volumi di dati da client inattivi.
- Non fornisci l'accesso allo storage su piattaforme e sistemi.

Vantaggi dell'adozione di questa best practice: usare file system o archiviazioni condivisi consente di distribuire i dati a uno o più consumatori senza doverli copiare. Questo consente di ridurre le risorse di archiviazione necessarie per il carico di lavoro.

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

Guida all'implementazione

Se hai più utenti o applicazioni che accedono agli stessi set di dati, usare una tecnologia di archiviazione condivisa è fondamentale per usare un'infrastruttura efficiente per il tuo carico di lavoro. La tecnologia di archiviazione condivisa offre una posizione centrale per archiviare e gestire set di dati ed evitare la loro duplicazione. Verifica anche la coerenza dei dati su sistemi diversi. Inoltre, la tecnologia di archiviazione condivisa consente un uso più efficiente della potenza di elaborazione, poiché più risorse di calcolo possono accedere ed elaborare i dati allo stesso momento in parallelo.

Acquisisci i dati dai servizi di archiviazione condivisa in base alle necessità e scollega i volumi non utilizzati per liberare le risorse.

Passaggi dell'implementazione

- Esegui la migrazione dei dati nell'archiviazione condivisa quando i dati hanno più consumer. Ecco alcuni esempi della tecnologia di archiviazione condivisa su AWS:

Storage option	When to use
Amazon EBS Multi-Attach	Amazon EBS Multi-Attach consente di collegare un volume singolo Provisioned IOPS SSD (io1 o io2) a più istanze che si trovano nella stessa zona di disponibilità.
Amazon EFS	See Quando scegliere Amazon EFS .
Amazon FSx	See Scegliere un Amazon FSx File System .
Amazon S3	Le applicazioni che non richiedono una struttura di file system e sono progettate per lavorare con lo storage degli oggetti possono usare Amazon S3 come soluzione di archiviazione degli oggetti a basso costo, durevole e altamente scalabile.

- Copia o acquisisci i dati solo da file system condivisi in base alle necessità. Come esempio, puoi creare un [file system Amazon FSx for Lustre supportato da Amazon S3](#) e caricare solo il sottoinsieme di dati richiesti per l'elaborazione dei processi su Amazon FSx.

- Elimina i dati nella modalità corretta per i tuoi modelli di utilizzo come definito in [SUS04-BP03 Utilizzo delle policy per gestire il ciclo di vita dei set di dati](#).
- Distacca i volumi dai client che non li utilizzano attivamente.

Risorse

Documenti correlati:

- [Collegare il file system a un bucket Amazon S3](#)
- [Usare Amazon EFS per AWS Lambda nelle applicazioni serverless](#)
- [Amazon EFS Intelligent-Tiering ottimizza i costi per carichi di lavoro con modelli di accesso mutevoli](#)
- [Uso di Amazon FSx con repository di dati on-premise](#)

Video correlati:

- [Storage cost optimization with Amazon EFS](#)
- [AWS re:Invent 2023 - What's new with AWS file storage](#)
- [AWS re:Invent 2023 - File storage for builders and data scientists on Amazon Elastic File System](#)

SUS04-BP07 Riduzione al minimo dello spostamento di dati tra reti

Usa file system condivisi o lo storage a oggetti per accedere ai dati comuni e contenere le risorse di rete totali necessarie per supportare i trasferimenti dei dati per il carico di lavoro.

Anti-pattern comuni:

- Archivi tutti i dati nella stessa Regione AWS, indipendentemente dalla posizione degli utenti.
- Non ottimizzi la dimensione e il formato dei dati prima di trasferirli sulla rete.

Vantaggi dell'adozione di questa best practice: l'ottimizzazione del trasferimento dei dati sulla rete riduce la quantità di risorse di rete totali richieste per il carico di lavoro e diminuisce l'impatto ambientale.

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

Guida all'implementazione

Trasferire i dati all'interno dell'organizzazione significa disporre di risorse di elaborazione, rete e archiviazione. Usa tecniche per ridurre il movimento dei dati e migliorare l'efficienza generale del tuo carico di lavoro.

Passaggi dell'implementazione

- Considera la vicinanza ai dati o agli utenti come un fattore importante nella fase decisionale per la [selezione di un'area geografica per il tuo carico di lavoro](#).
- Esegui la partizione dei servizi consumati a livello regionale in modo che i dati specifici della Regione siano archiviati nella Regione in cui sono usati.
- Usa formati di file efficienti (come Parquet oppure ORC) e comprimi i dati prima di spostarli sulla rete.
- Non trasferire dati inutilizzati. Alcuni esempi che possono aiutarti a evitare di spostare dati inutilizzati:
 - Riduci le risposte API solo ai dati pertinenti.
 - Aggrega i dati laddove richiesto (le informazioni a livello di record non sono necessarie).
 - Consulta [Well-Architected Lab - Optimize Data Pattern Using Amazon Redshift Data Sharing](#).
 - Consulta [Cross-account data sharing in AWS Lake Formation](#).
- Utilizza servizi in grado di supportarti nell'esecuzione del codice in posizioni più vicine agli utenti del carico di lavoro.

Service	When to use
Lambda@Edge	Usa per operazioni a uso intensivo di risorse di calcolo eseguite quando gli oggetti non si trovano nella cache.
Funzioni CloudFront	Usa per casi d'uso semplici, ad esempio manipolazioni di risposte o richieste HTTP(s) che possono essere avviate da funzioni di breve durata.

Service	When to use
AWS IoT Greengrass	Eseguire la memorizzazione nella cache di risorse di calcolo, messaggistica e dati per i dispositivi connessi.

Risorse

Documenti correlati:

- [Ottimizzazione dell'infrastruttura AWS per la sostenibilità, parte III: reti](#)
- [Infrastruttura globale AWS](#)
- [Funzionalità principali di Amazon CloudFront, tra cui CloudFront Global Edge Network](#)
- [Compressione delle richieste HTTP in Amazon OpenSearch Service](#)
- [Compressione intermedia dei dati con Amazon EMR](#)
- [Caricamento di file di dati compressi da Amazon S3 a Amazon Redshift](#)
- [Distribuzione dei file compressi con Amazon CloudFront](#)

Video correlati:

- [Demystifying data transfer on AWS](#)

Esempi correlati:

- [Architecting for sustainability - Minimize data movement across networks](#)

SUS04-BP08 Backup dei dati solo quando sono difficili da ricreare

Evita il back-up di dati senza valore aziendale per ridurre i requisiti delle risorse di archiviazione per il tuo carico di lavoro.

Anti-pattern comuni:

- Non hai una strategia di back-up per i tuoi dati.
- Esegui il back-up di dati che possono essere facilmente ricreati.

Vantaggi dell'adozione di questa best practice: se si evita il back-up di dati non critici si riduce la quantità di risorse di archiviazione richiesta per il carico di lavoro e si diminuisce l'impatto ambientale.

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

Guida all'implementazione

Evitando il back-up di dati non necessari si possono ridurre i costi e le risorse di archiviazione utilizzate dal carico di lavoro. Esegui il backup solo dei dati che hanno un valore aziendale o sono considerati necessari per soddisfare i requisiti di conformità. Esamina le policy di backup ed escludi l'archiviazione temporanea che non offre valore in uno scenario di ripristino.

Passaggi dell'implementazione

- Implementazione di una policy di classificazione dei dati come definito in [SUS04-BP01 Implementazione di una policy di classificazione dei dati](#).
- Usa le criticità della classificazione dei tuoi dati e la strategia di backup della progettazione basate su [Obiettivo del tempo di ripristino \(RTO\) e Obiettivo del punto di ripristino \(RPO\)](#). Evita il back-up di dati non critici.
 - Escludi i dati che possono essere facilmente ricreati.
 - Escludi dati temporanei dai backup.
 - Escludi copie locali dei dati, a meno che il tempo necessario per ripristinare tali dati da una posizione comune superi gli accordi sul livello di servizio (SLA).
- Usa una soluzione automatizzata o un servizio gestito per eseguire il back-up di dati aziendali strategici.
 - [AWS Backup](#) è un servizio completamente gestito che semplifica la centralizzazione e l'automazione della protezione dei dati tra i servizi AWS, nel cloud e on-premise. Per linee guida pratiche su come creare back-up automatizzati con AWS Backup, consulta [Well-Architected Labs: Test di backup e ripristino dei dati](#).
 - [Automatizza i back-up e ottimizza i costi di back-up per Amazon EFS con AWS Backup](#).

Risorse

Best practice correlate:

- [REL09-BP01 Identificazione e backup di tutti i dati che richiedono un backup o riproduzione dei dati dalle origini](#)

- [REL09-BP03 Esecuzione del backup dei dati in automatico](#)
- [REL13-BP02 Utilizzo di strategie di ripristino definite per conseguire gli obiettivi di ripristino](#)

Documenti correlati:

- [Utilizzo di AWS Backup per il backup e il ripristino dei file system di Amazon EFS](#)
- [Snapshot di Amazon EBS](#)
- [Lavorare con i backup su Amazon Relational Database Service](#)
- [Partner APN: partner per il backup](#)
- [Marketplace AWS: prodotti che possono essere utilizzati per il backup](#)
- [Backup di Amazon EFS](#)
- [Backup Amazon FSx per Windows File Server](#)
- [Backup e ripristino per Amazon ElastiCache for Redis](#)

Video correlati:

- [AWS re:Invent 2023 - Backup and disaster recovery strategies for increased resilience](#)
- [AWS re:Invent 2023 - What's new with AWS Backup](#)
- [AWS re:Invent 2021 - Backup, disaster recovery, and ransomware protection with AWS](#)

Esempi correlati:

- [Well-Architected Lab - Backup data](#)

Hardware e servizi

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

Best practice

- [SUS05-BP01 Utilizzo della quantità minima di hardware per soddisfare le esigenze aziendali](#)
- [SUS05-BP02 Utilizzo di tipi di istanze con il minimo impatto](#)

- [SUS05-BP03 Utilizzo dei servizi gestiti](#)
- [SUS05-BP04 Ottimizzazione dell'uso degli acceleratori di calcolo basati su hardware](#)

SUS05-BP01 Utilizzo della quantità minima di hardware per soddisfare le esigenze aziendali

Usa la quantità minima di hardware per il tuo carico di lavoro per soddisfare in modo efficiente le tue esigenze aziendali.

Anti-pattern comuni:

- Non monitori l'utilizzo delle risorse.
- Nella tua architettura sono presenti risorse con un basso livello di utilizzo.
- Non analizzi l'uso di hardware statico per stabilire se deve essere ridimensionato.
- Non imposti obiettivi di utilizzo dell'hardware per la tua infrastruttura di elaborazione in base a KPI aziendali.

Vantaggi dell'adozione di questa best practice: il corretto dimensionamento delle risorse cloud consente di ridurre l'impatto ambientale di un carico di lavoro, risparmiare denaro ed essere in linea con i benchmark di performance.

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

Guida all'implementazione

Seleziona con precisione la quantità di hardware richiesta dal tuo carico di lavoro per migliorare l'efficienza generale. Il Cloud AWS offre la flessibilità necessaria per espandere o ridurre il numero di risorse in modo dinamico attraverso una serie di meccanismi, come [AWS Auto Scaling](#) e soddisfare i cambiamenti della domanda. AWS offre anche [API e SDK](#) che consentono alle risorse di essere modificate con il minimo sforzo. Usa queste funzionalità per apportare modifiche frequenti alle implementazioni dei carichi di lavoro. Usa inoltre le linee guida sul dimensionamento corretto degli strumenti AWS per gestire le risorse cloud in modo efficiente e soddisfare le esigenze aziendali.

Passaggi dell'implementazione

- Scegli il tipo di istanza: scegli il tipo di istanza giusto per soddisfare al meglio le tue esigenze. Per informazioni su come scegliere le istanze Amazon Elastic Compute Cloud e utilizzare meccanismi quali la selezione delle istanze basata sugli attributi, consulta le seguenti risorse:

- [Come faccio a scegliere il tipo di istanza Amazon EC2 appropriato per il mio carico di lavoro?](#)
- [Selezione del tipo di istanza basata sugli attributi per il parco istanze Amazon EC2.](#)
- [Create an Auto Scaling group using attribute-based instance type selection.](#)
- Scala: utilizza piccoli incrementi per scalare i carichi di lavoro variabili.
- Utilizza più opzioni di acquisto di calcolo: bilancia flessibilità, scalabilità e risparmio sui costi delle istanze con più opzioni di acquisto di calcolo.
 - Le [istanze on-demand Amazon EC2](#) sono più adatte a carichi di lavoro nuovi, stateful e con picchi che non possono essere flessibili dal punto di vista dei tipi di istanza, della posizione o dal punto di vista temporale.
 - Le [istanze spot Amazon EC2](#) sono un ottimo modo per fornire altre opzioni alle applicazioni flessibili e tolleranti ai guasti.
 - Sfrutta i [Savings Plans per il calcolo](#) per carichi di lavoro stazionari che consentono soluzioni flessibili se le tue esigenze (come AZ, area geografica, famiglie di istanze o tipi di istanze) cambiano.
- Usa la diversità di istanze e zone di disponibilità: massimizza la disponibilità delle applicazioni e sfrutta la capacità in eccesso diversificando le istanze e le zone di disponibilità.
- Dimensiona correttamente le istanze: utilizza i suggerimenti per il dimensionamento corretto degli strumenti AWS per regolare il carico di lavoro. Per ulteriori informazioni, consulta [Optimizing your cost with Rightsizing Recommendations](#) e [Right Sizing: Provisioning Instances to Match Workloads](#)
 - Puoi utilizzare i suggerimenti per il dimensionamento corretto in AWS Cost Explorer o [AWS Compute Optimizer](#) per identificare le corrette opportunità di dimensionamento.
- Negozia gli accordi sul livello di servizio (service-level agreement, SLA): negozia gli SLA che consentono di ridurre temporaneamente la capacità mentre l'automazione implementa le risorse sostitutive.

Risorse

Documenti correlati:

- [Ottimizzazione dell'infrastruttura AWS per la sostenibilità, Parte I: elaborazione](#)
- [Selezione del tipo di istanza basata sugli attributi per Auto Scaling per il parco istanze Amazon EC2](#)
- [Documentazione di AWS Compute Optimizer](#)
- [Uso di Lambda: ottimizzazione delle performance](#)
- [Documentazione sulla scalabilità automatica](#)

Video correlati:

- [AWS re:Invent 2023 - What's new with Amazon EC2](#)
- [AWS re:Invent 2023 - Smart savings: Amazon Elastic Compute Cloud cost-optimization strategies](#)
- [AWS re:Invent 2022 - Optimizing Amazon Elastic Kubernetes Service for performance and cost on AWS](#)
- [AWS re:Invent 2023 - Sustainable compute: reducing costs and carbon emissions with AWS](#)

SUS05-BP02 Utilizzo di tipi di istanze con il minimo impatto

Esegui un monitoraggio costante e usa nuovi tipi di istanza per sfruttare le migliorie in termini di efficienza energetica.

Anti-pattern comuni:

- Utilizzi una sola famiglia di istanze.
- Utilizzi solo istanze x86.
- Specifichi un tipo di istanza nella configurazione Amazon EC2 Auto Scaling.
- Utilizzi istanze AWS in un modo per il quale non sono state progettate, ad esempio utilizzi istanze ottimizzate per il calcolo per un carico di lavoro a uso intensivo della memoria.
- Non valuti regolarmente l'uso di nuovi tipi di istanza.
- Non segui i consigli ricevuti dagli strumenti di dimensionamento AWS, ad esempio [AWS Compute Optimizer](#).

Vantaggi dell'adozione di questa best practice: l'uso di istanze energeticamente efficienti e di dimensioni corrette consente di ridurre in modo considerevole l'impatto ambientale e i costi del carico di lavoro.

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

Guida all'implementazione

L'uso di istanze efficienti nel carico di lavoro cloud è fondamentale per ridurre l'utilizzo delle risorse e i costi. Monitora costantemente il rilascio di nuovi tipi di istanza e sfrutta le migliorie in tema di efficienza energetica, inclusi i tipi di istanza progettati per supportare carichi di lavoro specifici, come la formazione del machine learning, le inferenze e la transcodifica dei video.

Passaggi dell'implementazione

- Scopri e approfondisci i tipi di istanze: individua i tipi di istanze che possono ridurre l'impatto ambientale del tuo carico di lavoro.
 - Iscriviti alle [Novità di AWS](#) per rimanere aggiornato sulle più recenti tecnologie e istanze AWS.
 - Approfondisci i vari tipi di istanza AWS.
 - Scopri di più sulle istanze basate su AWS Graviton che offrono le performance migliori performance per watt di energia utilizzata in Amazon EC2 guardando [re:Invent 2020 - Approfondimento sulle istanze Amazon EC2 alimentate dal processore AWS Graviton2](#) e [Approfondimento sulle istanze AWS Graviton3 e Amazon EC2 C7g](#).
- Usa i tipi di istanza che comportano il minor impatto: pianifica la transizione del carico di lavoro ai tipi di istanza con il minor impatto.
 - Definisci un processo per valutare nuove funzionalità o istanze per il carico di lavoro. Sfrutta l'agilità del cloud per testare in modo semplice e rapido in che modo i nuovi tipi di istanza possono migliorare la sostenibilità ambientale del carico di lavoro. Utilizza metriche proxy per misurare la quantità di risorse necessarie per completare un'unità di lavoro.
 - Se possibile, modifica il carico di lavoro in modo che funzioni con diversi numeri di CPU e quantità di memoria diverse per massimizzare la scelta del tipo di istanza.
 - Valuta l'ipotesi di trasferire il carico di lavoro in istanze basate su Graviton per migliorare l'efficienza delle prestazioni del carico di lavoro. Per ulteriori informazioni sullo spostamento dei carichi di lavoro su AWS Graviton, consulta [AWS Graviton Fast Start](#) e [Considerations when transitioning workloads to AWS Graviton-based Amazon Elastic Compute Cloud instances](#).
 - Valuta l'ipotesi di selezionare l'opzione AWS Graviton quando utilizzi i [servizi gestiti AWS](#).
 - Esegui la migrazione del carico di lavoro nelle regioni che offrono istanze con il minor impatto in termini di sostenibilità e che contemporaneamente soddisfano i requisiti aziendali.
 - Per i carichi di lavoro di machine learning, sfrutta l'hardware specifico per il tuo carico di lavoro, come ad esempio [AWS Trainium](#), [AWS Inferentia](#) e [Amazon EC2 DL1](#). Le istanze AWS Inferentia come Inf2 offrono fino al 50% in più di prestazioni per watt rispetto alle istanze Amazon EC2 paragonabili.
 - Usa [Amazon SageMaker Inference Recommender](#) per dimensionare correttamente l'endpoint dell'inferenza ML.
 - Per carichi di lavoro con picchi (carichi di lavoro con requisiti non frequenti di capacità aggiuntiva), utilizza [istanze a prestazioni espandibili](#).

- Usa le [Istanze spot Amazon EC2](#) per carichi di lavoro stateless e con tolleranza ai guasti per aumentare l'utilizzo complessivo del cloud e ridurre l'impatto di sostenibilità delle risorse inutilizzate.
- Esegui e ottimizza: esegui e ottimizza l'istanza del carico di lavoro.
- Per carichi di lavoro effimeri valuta le [metriche Amazon CloudWatch dell'istanza](#), come CPUUtilization per verificare se l'istanza è inattiva o sottoutilizzata.
- Per i carichi di lavoro stabili, esegui controlli con gli strumenti di ridimensionamento AWS, come [AWS Compute Optimizer](#) a intervalli regolari per individuare eventuali opportunità di ottimizzazione e ridimensionamento corretto delle istanze.
- [Well-Architected Lab: Raccomandazioni per il dimensionamento appropriato](#)
- [Well-Architected Lab: Ridimensionamento con Compute Optimizer](#)
- [Well-Architected Lab: ottimizzazione dei modelli hardware e conformità agli indicatori KPI di sostenibilità](#)

Risorse

Documenti correlati:

- [Ottimizzazione dell'infrastruttura AWS per la sostenibilità, Parte I: elaborazione](#)
- [AWS Graviton](#)
- [Amazon EC2 DL1](#)
- [Parchi istanza di prenotazione della capacità Amazon EC2](#)
- [Serie di istanze spot Amazon EC2](#)
- [Funzioni: configurazione della funzione Lambda](#)
- [Attribute-based instance type selection for Amazon EC2 Fleet](#)
- [Building Sustainable, Efficient, and Cost-Optimized Applications on AWS](#)
- [How the Contino Sustainability Dashboard Helps Customers Optimize Their Carbon Footprint](#)

Video correlati:

- [AWS re:Invent 2023 - AWS Graviton: The best price performance for your AWS workloads](#)
- [AWS re:Invent 2023 - New Amazon Elastic Compute Cloud generative AI capabilities in AWS Management Console](#)

- [AWS re:Invent 2023 - What's new with Amazon Elastic Compute Cloud](#)
- [AWS re:Invent 2023 - Smart savings: Amazon Elastic Compute Cloud cost-optimization strategies](#)
- [AWS re:Invent 2021 - Deep dive into AWS Graviton3 and Amazon EC2 C7g instances](#)
- [AWS re:Invent 2022 - Build a cost-, energy-, and resource-efficient compute environment](#)

Esempi correlati:

- [Solution: Guidance for Optimizing Deep Learning Workloads for Sustainability on AWS](#)
- [Migrating Amazon Relational Database Service Databases to Graviton](#)

SUS05-BP03 Utilizzo dei servizi gestiti

Usa i servizi gestiti per operare in modo più efficiente nel cloud.

Anti-pattern comuni:

- Usi istanze Amazon EC2 in modo ridotto per eseguire le tue applicazioni.
- Il tuo team interno gestisce solo il carico di lavoro, senza tempo per focalizzarsi sull'innovazione o sulle semplificazioni.
- Implementi e mantieni tecnologie per attività che possono essere eseguite in modo più efficiente sui servizi gestiti.

Vantaggi dell'adozione di questa best practice:

- L'uso dei servizi gestiti sposta la responsabilità su AWS, che ha visibilità su milioni di clienti, i quali possono contribuire alla promozione di nuove innovazioni ed efficienze.
- Il servizio gestito distribuisce l'impatto ambientale del servizio su molti utenti a causa dei piani di controllo multi-tenant.

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

Guida all'implementazione

I servizi gestiti consentono di affidare ad AWS la responsabilità di mantenere un utilizzo alto e un'ottimizzazione della sostenibilità dell'hardware implementato. I servizi gestiti eliminano anche

l'onere operativo e amministrativo legato alla manutenzione di un servizio, consentendo al tuo team di avere più tempo e di concentrarsi sull'innovazione.

Esamina il carico di lavoro per identificare i componenti che possono essere sostituiti dai servizi gestiti AWS. Ad esempio, [Amazon RDS](#), [Amazon Redshift](#) e [Amazon ElastiCache](#) offrono un servizio di database gestito. [Amazon Athena](#), [Amazon EMR](#) e [Amazon OpenSearch Service](#) offrono un servizio di analisi gestito.

Passaggi dell'implementazione

1. Esegui l'inventario del carico di lavoro: esegui l'inventario di servizi e componenti del tuo carico di lavoro.
2. Identifica i candidati: valuta e identifica i componenti che possono essere sostituiti dai servizi gestiti. Ecco alcuni esempi in cui potresti prendere in considerazione l'uso di un servizio gestito:

Task	What to use on AWS
Ospitare un database	Usa istanze gestite Amazon Relational Database Service (Amazon RDS) invece di mantenere le tue istanze Amazon RDS su Amazon Elastic Compute Cloud (Amazon EC2) .
Ospitare il carico di lavoro di un container	Usa AWS Fargate , invece di implementare un'infrastruttura di container proprietaria.
Ospitare applicazioni Web	Usa l' Hosting AWS Amplify come CI/CD completamente gestito e servizio di hosting per siti Web statici e app Web con rendering lato server.

3. Crea un piano di migrazione: identifica le dipendenze e crea un piano di migrazione. Aggiorna runbook e playbook.
 - [AWS Application Discovery Service](#) raccoglie e illustra automaticamente informazioni dettagliate sulle dipendenze delle applicazioni e sul loro utilizzo per aiutarti a prendere decisioni più informate durante la pianificazione della migrazione.
4. Esegui i test: esegui i test prima di migrare al servizio gestito.

5. Sostituisci i servizi self-hosted: usa il piano di migrazione per sostituire i servizi self-hosted con servizi gestiti.
6. Monitora e modifica: monitora costantemente il servizio al termine della migrazione per apportare le modifiche richieste e ottimizzare il servizio.

Risorse

Documenti correlati:

- [Prodotti Cloud AWS](#)
- [Calcolatore del costo totale di proprietà \(TCO\) di AWS](#)
- [Amazon DocumentDB](#)
- [Amazon Elastic Kubernetes Service \(EKS\)](#)
- [Amazon Managed Streaming for Apache Kafka \(Amazon MSK\)](#)

Video correlati:

- [AWS re:Invent 2021 - Cloud operations at scale with AWS Managed Services](#)
- [AWS re:Invent 2023 - Best practices for operating on AWS](#)

SUS05-BP04 Ottimizzazione dell'uso degli acceleratori di calcolo basati su hardware

Ottimizza l'uso delle istanze a calcolo accelerato per ridurre i requisiti dell'infrastruttura fisica del carico di lavoro.

Anti-pattern comuni:

- Utilizzo delle GPU non monitorato.
- Utilizzo di un'istanza generica per il carico di lavoro quando un'istanza appositamente sviluppata potrebbe offrire prestazioni più elevate, costi inferiori e migliori prestazioni per watt.
- Utilizzo di acceleratori di calcolo basati su hardware per attività in cui sono più efficienti le alternative basate su CPU.

Vantaggi dell'adozione di questa best practice: ottimizzando l'uso degli acceleratori basati su hardware, è possibile ridurre le richieste di infrastruttura fisica del carico di lavoro.

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

Guida all'implementazione

Se si necessita di un'elevata capacità di elaborazione, si può trarre vantaggio dall'uso di istanze a calcolo accelerato, che forniscono l'accesso ad acceleratori di calcolo basati su hardware, come le unità di elaborazione grafica (GPU) e i field programmable gate array (FPGA). Questi acceleratori hardware eseguono alcune funzioni, come l'elaborazione grafica o la rilevazione della corrispondenza dei modelli di dati, in modo più efficiente rispetto alle alternative basate su CPU. Molti carichi di lavoro accelerati, come il rendering grafico, la transcodifica e il machine learning, sono altamente variabili in termini di utilizzo di risorse. Mantieni in esecuzione questo tipo di hardware solo per il tempo necessario e disattivalo automaticamente quando non serve per ridurre la quantità di risorse utilizzate.

Passaggi dell'implementazione

- Identifica quali [istanze a calcolo accelerato](#) possono soddisfare i tuoi requisiti.
- Per i carichi di lavoro di machine learning, sfrutta l'hardware specifico per il tuo carico di lavoro, come ad esempio [AWS Trainium](#), [AWS Inferentia](#) e [Amazon EC2 DL1](#). Le istanze AWS Inferentia come Inf2 offrono fino al [50% in più di prestazioni per watt rispetto alle istanze Amazon EC2 paragonabili](#).
- Raccogli i parametri di utilizzo delle istanze a calcolo accelerato. Ad esempio, puoi utilizzare l'agente CloudWatch per raccogliere parametri come `utilization_gpu` e `utilization_memory` per le GPU come mostrato in [Raccolta dei parametri delle GPU NVIDIA con Amazon CloudWatch](#).
- Ottimizza il codice, il funzionamento della rete e le impostazioni degli acceleratori hardware per garantire il pieno utilizzo dell'hardware sottostante.
 - [Ottimizza l'impostazioni delle GPU](#)
 - [Monitoraggio e ottimizzazione delle GPU nell'AMI per il Deep Learning](#)
 - [Ottimizzazione dell'I/O per la messa a punto delle prestazioni delle GPU dedicate all'addestramento del deep learning in Amazon SageMaker](#)
- Utilizzate le librerie e i driver per GPU più recenti e performanti.
- Utilizza l'automazione per rilasciare le istanze GPU non in uso.

Risorse

Documenti correlati:

- [Calcolo accelerato](#)
- [Progettiamo! Sviluppo di architetture con chip e acceleratori personalizzati](#)
- [Come faccio a scegliere il tipo di istanza Amazon EC2 appropriato per il mio carico di lavoro?](#)
- [Istanze Amazon EC2 VT1](#)
- [Scegli il miglior acceleratore IA e compilatore del modello per l'inferenza nella visione artificiale con Amazon SageMaker](#)

Video correlati:

- [AWS re:Invent 2021 - How to select Amazon EC2 GPU instances for deep learning](#)
- [AWS Online Tech Talks - Deploying Cost-Effective Deep Learning Inference](#)
- [AWS re:Invent 2023 - Cutting-edge AI with AWS and NVIDIA](#)
- [AWS re:Invent 2022 - \[NEW LAUNCH!\] Introducing AWS Inferentia2-based Amazon EC2 Inf2 instances](#)
- [AWS re:Invent 2022 - Accelerate deep learning and innovate faster with AWS Trainium](#)
- [AWS re:Invent 2022 - Deep learning on AWS with NVIDIA: From training to deployment](#)

Processo e cultura

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

Best practice

- [SUS06-BP01 Adozione di metodi che consentano di introdurre rapidamente migliorie in tema di sostenibilità](#)
- [SUS06-BP02 Aggiornamento del carico di lavoro](#)
- [SUS06-BP03 Aumento dell'utilizzo degli ambienti di costruzione](#)
- [SUS06-BP04 Utilizzo di device farm gestite per i test](#)

SUS06-BP01 Adozione di metodi che consentano di introdurre rapidamente migliorie in tema di sostenibilità

Adotta metodi e processi per convalidare migliorie potenziali, ridurre i costi legati ai test e offrire piccole migliorie.

Anti-pattern comuni:

- Analizzare l'applicazione rispetto alla sostenibilità è un'attività che viene eseguita solo una volta, all'inizio di un progetto.
- Il tuo carico di lavoro non è aggiornato, poiché il processo di rilascio è troppo complesso per introdurre modifiche minori per l'efficienza delle risorse.
- Non hai meccanismi per migliorare il tuo carico di lavoro in termini di sostenibilità.

Vantaggi dell'adozione di questa best practice: definendo un processo per avviare e monitorare le migliorie in termini di sostenibilità, potrai adottare continuamente nuove funzionalità, eliminare i problemi e migliorare l'efficienza del carico di lavoro.

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

Guida all'implementazione

Testa e convalida potenziali miglioramenti all'impatto sulla sostenibilità prima di implementarli in produzione. Tieni in considerazione il costo dei test quando calcoli il potenziale vantaggio futuro di un miglioramento. Sviluppa metodi di test a basso costo per consentire la distribuzione di piccoli miglioramenti.

Passaggi dell'implementazione

- Comprendi e comunica gli obiettivi di sostenibilità organizzativa: comprendi gli obiettivi di sostenibilità organizzativa, come la riduzione delle emissioni di carbonio o la gestione delle risorse idriche. Traduci questi obiettivi in requisiti di sostenibilità per i carichi di lavoro del cloud. Comunica questi requisiti alle principali parti interessate.
- Aggiungi i requisiti di sostenibilità al backlog: aggiungi i requisiti per il miglioramento della sostenibilità al backlog di sviluppo.
- Itera e migliora: usa un [processo di miglioramento iterativo](#) che ti consente di identificare, valutare, dare la priorità, testare e implementare questi miglioramenti.

- Esegui il test utilizzando il prodotto minimo funzionante (MVP): sviluppa e testa i potenziali miglioramenti utilizzando i componenti rappresentativi del prodotto minimo funzionante per ridurre i costi e l'impatto ambientale dei test.
- Semplifica il processo: migliora e semplifica costantemente i processi di sviluppo. Ad esempio, automatizza il processo di distribuzione del software con pipeline di distribuzione e integrazione continue (CI/CD) per testare e implementare migliorie potenziali per ridurre il livello di impegno e gli errori causati da processi manuali.
- Gestisci formazione e sensibilizzazione: gestisci i programmi di formazione per i membri del team per istruirli sulla sostenibilità e sull'impatto delle attività rispetto agli obiettivi di sostenibilità organizzativa.
- Valuta e modifica: valuta continuamente l'impatto dei miglioramenti e apporta le modifiche necessarie.

Risorse

Documenti correlati:

- [AWS offre soluzioni di sostenibilità](#)
- [Procedure di sviluppo agile e scalabile basate su AWS CodeCommit](#)

Video correlati:

- [AWS re:Invent 2023 - Sustainable architecture: Past, present, and future](#)
- [AWS re:Invent 2022 - Delivering sustainable, high-performing architectures](#)
- [AWS re:Invent 2022 - Architecting sustainably and reducing your AWS carbon footprint](#)
- [AWS re:Invent 2022 - Sustainability in AWS global infrastructure](#)
- [AWS re:Invent 2023 - What's new with AWS observability and operations](#)

Esempi correlati:

- [Well-Architected Lab - Trasformare i report su costi e utilizzo in report sull'efficienza](#)

SUS06-BP02 Aggiornamento del carico di lavoro

Aggiorna il tuo carico di lavoro per adottare funzionalità efficienti, eliminare le problematiche e migliorare l'efficienza generale del tuo carico di lavoro.

Anti-pattern comuni:

- Ritieni che l'architettura corrente diventi statica e non venga aggiornata nel corso del tempo.
- Non disponi di sistemi né esegui regolarmente una valutazione per la compatibilità di software e pacchetti aggiornati con il carico di lavoro.

Vantaggi dell'adozione di questa best practice: la definizione di un processo per garantire il costante aggiornamento del carico di lavoro ti consentirà di adottare nuove caratteristiche e funzionalità, risolvere i problemi e migliorare l'efficienza del carico di lavoro.

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

Guida all'implementazione

Sistemi operativi, runtime, middleware, librerie e applicazioni aggiornati possono incidere sull'efficienza dei carichi di lavoro e facilitano l'adozione delle tecnologie più efficienti. Il software aggiornato potrebbe anche includere funzionalità per misurare in modo più accurato l'impatto in termini di sostenibilità del carico di lavoro, poiché i fornitori offrono caratteristiche per raggiungere i propri obiettivi di sostenibilità. Adotta una cadenza regolare per aggiornare il tuo carico di lavoro con le ultime funzionalità e i rilasci più recenti.

Passaggi dell'implementazione

- Definisci un processo: usa un processo e una pianificazione per valutare nuove funzionalità o istanze per il carico di lavoro. Sfrutta l'agilità del cloud per testare in modo semplice e rapido il modo in cui le nuove funzionalità possono migliorare il carico di lavoro nei seguenti ambiti:
 - Riduzione dell'impatto a livello di sostenibilità.
 - Raggiungimento di maggiore efficienza in termini di prestazioni.
 - Eliminazione delle barriere finalizzata a un miglioramento pianificato.
 - Miglioramento della capacità di misurare e gestire l'impatto a livello di sostenibilità.
- Esegui l'inventario: redigi l'inventario del software e dell'architettura del carico di lavoro e identifica i componenti che richiedono un aggiornamento.

- Puoi usare [AWS Systems Manager Inventory](#) per raccogliere i metadati relativi a sistema operativo (SO), applicazioni e istanze dalle istanze Amazon EC2 per avere una panoramica immediata su quali istanze stanno eseguendo il software e le configurazioni richieste dalle policy software e quali istanze devono essere aggiornate.
- Apprendi la procedura di aggiornamento: scopri come aggiornare i componenti del carico di lavoro.

Workload component	How to update
Immagini della macchina	Usa EC2 Image Builder per gestire gli aggiornamenti alle Amazon Machine Images (AMI) per le immagini Linux o Windows Server.
Immagini del container	Usa Amazon Elastic Container Registry (Amazon ECR) con la pipeline esistente per gestire le immagini Amazon Elastic Container Service (Amazon ECS) .
AWS Lambda	AWS Lambda include funzionalità di gestione delle versioni .

- Utilizza l'automazione: usa l'automazione degli aggiornamenti per ridurre il livello di impegno per implementare le nuove funzionalità e limitare gli errori causati dai processi manuali.
- Puoi usare [CI/CD](#) per aggiornare automaticamente le AMI, le immagini di container e altri artefatti relativi alla tua applicazione cloud.
- Puoi usare strumenti come [AWS Systems Manager Patch Manager](#) per automatizzare il processo degli aggiornamenti di sistema e pianificare le attività tramite [Finestre di manutenzione AWS Systems Manager](#).

Risorse

Documenti correlati:

- [Centro di progettazione AWS](#)
- [Le novità di AWS](#)
- [Strumenti per sviluppatori in AWS](#)

Video correlati:

- [AWS re:Invent 2022 - Optimize your AWS workloads with best-practice guidance](#)
- [All Things Patch: AWS Systems Manager](#)

Esempi correlati:

- [Well-Architected Labs: Inventario e gestione delle patch](#)
- [Laboratorio: AWS Systems Manager](#)

SUS06-BP03 Aumento dell'utilizzo degli ambienti di costruzione

Aumenta l'uso delle risorse per sviluppare, testare e creare i tuoi carichi di lavoro.

Anti-pattern comuni:

- Esegui il provisioning manuale o interrompi i tuoi ambienti di sviluppo.
- Fai in modo che i tuoi ambienti di sviluppo siano in esecuzione indipendentemente dalle attività di test, creazione o rilascio (ad esempio, eseguire un ambiente al di fuori dell'orario di lavoro dei membri del tuo team di sviluppo).
- Esegui un provisioning eccessivo delle tue risorse per gli ambienti di creazione.

Vantaggi derivanti dall'adozione di questa best practice: aumentando l'uso degli ambienti di sviluppo, puoi migliorare l'efficienza complessiva del tuo carico di lavoro cloud, allocando al contempo le risorse di cui gli sviluppatori hanno bisogno per creare, testare e sviluppare in modo efficiente.

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

Guida all'implementazione

Utilizza l'automazione e l'infrastruttura come codice per rendere operativi gli ambienti di produzione quando necessario e dismetterli quando non vengono utilizzati. Un modello comune consiste nel pianificare periodi di disponibilità che coincidano con l'orario di lavoro dei membri del team incaricati dello sviluppo. Gli ambienti di test devono essere molto simili alla configurazione di produzione. Tuttavia, cerca la possibilità di utilizzare tipi di istanze con capacità di espansione, istanze Spot Amazon EC2, servizi di database con dimensionamento automatico, container e tecnologie serverless per allineare la capacità di sviluppo e test all'uso. Limita i volumi di dati per soddisfare solo

i requisiti di test. Se usi i dati di produzione per i test, rifletti sulla possibilità di condividere i dati di produzione invece di spostarli.

Passaggi dell'implementazione

- Usa l'infrastructure as code: usa l'infrastructure as code per eseguire il provisioning dei tuoi ambienti di compilazione.
- Usa l'automazione: utilizza l'automazione per gestire il ciclo di vita degli ambienti di sviluppo e test e massimizzare l'efficienza delle tue risorse di compilazione.
- Ottimizza l'utilizzo: usa le strategie per ottimizzare l'utilizzo degli ambienti di sviluppo e test.
 - Utilizza ambienti rappresentativi minimi realizzabili per lo sviluppo e il test di potenziali miglioramenti.
 - Utilizza tecnologie serverless, se possibile.
 - Utilizza istanze on-demand per integrare i dispositivi per gli sviluppatori.
 - Utilizza i tipi di istanze con capacità di espansione, istanze Spot e altre tecnologie per allineare la capacità di compilazione all'uso.
 - Adotta servizi cloud nativi per un accesso sicuro alle shell delle istanze invece di implementare parchi istanze di host bastion.
 - Dimensiona automaticamente le tue risorse di sviluppo in base alle tue attività.

Risorse

Documenti correlati:

- [AWS Systems Manager Session Manager](#)
- [Istanze espandibili di prestazioni Amazon EC2](#)
- [Che cos'è AWS CloudFormation?](#)
- [Che cos'è AWS CodeBuild?](#)
- [Instance Scheduler su AWS](#)

Video correlati:

- [AWS re:Invent 2023 - Continuous integration and delivery for AWS](#)

SUS06-BP04 Utilizzo di device farm gestite per i test

Usa device farm gestite per testare in maniera efficiente una nuova funzionalità su un set rappresentativo di hardware.

Anti-pattern comuni:

- Testa e distribuisce manualmente la tua applicazione su singoli dispositivi fisici.
- Non usare il servizio di test delle app per testare e interagire con le tue app (ad esempio, Android, iOS e app Web) su dispositivi fisici reali.

Vantaggi dell'adozione di questa best practice: usare le device farm gestite per testare applicazioni abilitate al cloud offre una serie di vantaggi:

- Offrono funzionalità più efficienti per testare le applicazioni su un'ampia gamma di dispositivi.
- Eliminano la necessità di un'infrastruttura in-house per i test.
- Offrono diverse tipologie di dispositivi, tra cui hardware di generazioni precedenti e meno diffuso, eliminando così la necessità di aggiornamenti non necessari dei dispositivi.

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

Guida all'implementazione

L'uso di device farm gestite può aiutarti a semplificare il processo di test per le nuove funzionalità su un gruppo rappresentativo di hardware. Le device farm gestite offrono diversi tipi di dispositivi, inclusi hardware meno diffusi e di generazioni precedenti, ed evitano l'impatto sulla sostenibilità dei clienti dovuti ad aggiornamenti dei dispositivi non necessari.

Passaggi dell'implementazione

- Definisci i requisiti di test: stabilisci i requisiti di test ed esegui la pianificazione, ad esempio il tipo di test, i sistemi operativi e il programma di test.
 - Puoi usare [Amazon CloudWatch RUM](#) per raccogliere e analizzare i dati lato client e formulare il tuo piano di test.
- Scegli una device farm gestita: seleziona una device farm gestita in grado di supportare i tuoi requisiti di test. Ad esempio, puoi usare [AWS Device Farm](#) per testare e comprendere l'impatto delle tue modifiche su un set rappresentativo di hardware.

- Usa l'automazione: utilizza automazione e integrazione continua/implementazione continua (CI/CD) per pianificare ed eseguire i test.
 - [Integrazione di Device Farm AWS con la pipeline CI/CD per eseguire i test Selenium sui diversi browser](#)
 - [Creazione e test di app iOS e iPadOS con AWS DevOps e servizi mobili](#)
- Rivedi e modifica: revisiona continuamente i risultati dei test e apporta le migliorie necessarie.

Risorse

Documenti correlati:

- [Elenco dei dispositivi AWS Device Farm](#)
- [Visualizzazione del dashboard CloudWatch RUM](#)

Esempi correlati:

- [App di esempio AWS Device Farm per Android](#)
- [App di esempio AWS Device Farm per iOS](#)
- [Test Appium Web per AWS Device Farm](#)

Video correlati:

- [AWS re:Invent 2023 - Improve your mobile and web app quality using AWS Device Farm](#)
- [AWS re:Invent 2021 - Optimize applications through end user insights with Amazon CloudWatch RUM](#)

Conclusione

Un numero crescente di organizzazioni definisce obiettivi di sostenibilità in risposta ai cambiamenti delle normative di governo, ai vantaggi competitivi e alla richiesta di clienti, dipendenti e investitori. CTO, progettisti, sviluppatori e membri dei team operativi sono alla ricerca di modalità per contribuire direttamente agli obiettivi di sostenibilità della propria organizzazione. Utilizzando questi principi e best practice di progettazione supportati dai servizi AWS, puoi prendere decisioni informate, bilanciando sicurezza, costi, performance, affidabilità ed eccellenza operativa con risultati di sostenibilità per i carichi di lavoro Cloud AWS. Ogni azione che intraprendi per ridurre l'utilizzo di risorse e aumentare l'efficienza nei carichi di lavoro contribuisce a una riduzione dell'impatto ambientale e agli obiettivi di sostenibilità più ampi delle organizzazioni.

Collaboratori

I collaboratori di questo documento includono:

- Sam Mokhtari, Senior Efficiency Lead Solutions Architect, Amazon Web Services
- Brendan Sisson, Principal Sustainability Solutions Architect, Amazon Web Services
- Margaret O'Toole, Sustainability Tech Leader, Amazon Web Services
- Steffen Grunwald, Principal Sustainability Solutions Architect, Amazon Web Services
- Ryan Eccles, Principal Engineer, Sustainability, Amazon
- Rodney Lester, Principal Architect, Amazon Web Services
- Adrian Cockcroft, VP Sustainability Architecture, Amazon Web Services
- Ian Meyers, Director of Technology, Solutions Architecture, Amazon Web Services

Approfondimenti

Per ulteriori informazioni, fare riferimento a:

- [AWS Well-Architected](#)
- [Centro di progettazione AWS](#)
- [Sostenibilità nel cloud](#)
- [AWS offre soluzioni di sostenibilità](#)
- [The Climate Pledge](#)
- [Obiettivi di sviluppo sostenibile delle Nazioni Unite](#)
- [Greenhouse Gas Protocol](#)

Revisioni del documento

Per ricevere una notifica sugli aggiornamenti di questo whitepaper, iscriviti al feed RSS.

Modifica	Descrizione	Data
Whitepaper aggiornato	Best practice aggiornate con nuova guida all'implementazione.	June 27, 2024
Livelli di rischio aggiornati	Aggiornamenti di minore entità ai livelli di rischio delle best practice.	October 3, 2023
Linee guida sulle best practice aggiornate	Le best practice sono state aggiornate con nuove linee guida nelle seguenti aree: Allineamento alla domanda , Software e architettura , Data Hardware e servizi .	July 13, 2023
Aggiornamenti per il nuovo framework	Best practice aggiornate con prontuario e nuove best practice aggiunte.	April 10, 2023
Whitepaper aggiornato	Best practice aggiornate con nuova guida all'implementazione.	December 15, 2022
Whitepaper aggiornato	Ampliamento delle best practice e aggiunta dei piani di miglioramento.	October 20, 2022
Pubblicazione originale	Pubblicazione del Principio della sostenibilità – Framework AWS Well-Architected.	December 2, 2021

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 di AWS vengono forniti "così come sono", senza garanzie, rappresentazioni o condizioni di nessun tipo, sia espresse che implicite. Le responsabilità e gli obblighi di AWS verso i propri clienti sono disciplinati dagli accordi AWS e il presente documento non fa parte né modifica alcun accordo tra AWS e i suoi clienti.

© 2021 Amazon Web Services, Inc. o sue affiliate. Tutti i diritti riservati.

AWS Glossary

For the latest AWS terminology, see the [AWS glossary](#) in the Glossario AWS Reference.