AWS Well-Architected Framework

Disaster recovery dei carichi di lavoro su AWS: Recovery nel cloud

Disaster recovery dei carichi di lavoro su AWS: Recovery nel cloud: AWS Well-Architected Framework

Table of Contents

Disaster recovery dei carichi di lavoro su AWS: Recovery nel cloud

Data di pubblicazione: 12 febbraio 2021 () Cronologia dei documenti

Il disaster recovery è il processo di preparazione e ripristino da un disastro. Un evento che impedisce a un carico di lavoro o a un sistema di raggiungere gli obiettivi aziendali nella sua sede principale di implementazione è considerato un disastro. Questo paper descrive le migliori pratiche per pianificare e testare il disaster recovery per qualsiasi carico di lavoro distribuito e offre diversi approcci per mitigare i rischi e soddisfare il Recovery Time Objective (RTO) e il Recovery Point Objective (RPO) per quel carico di lavoro. AWS

Questo white paper spiega come implementare il disaster recovery per i carichi di lavoro su. AWS Per informazioni sull'utilizzo AWS come sito di disaster recovery AWS per carichi di lavoro locali, consulta la sezione Disaster Recovery of On-Premises Applications.

Introduzione

Il carico di lavoro deve svolgere la funzione prevista in modo corretto e coerente. Per raggiungere questo obiettivo, è necessario progettare per la resilienza. La resilienza è la capacità di un carico di lavoro di riprendersi da interruzioni dell'infrastruttura, del servizio o delle applicazioni, acquisire dinamicamente risorse di elaborazione per soddisfare la domanda e mitigare le interruzioni, come configurazioni errate o problemi transitori di rete.

Il disaster recovery (DR) è una parte importante della strategia di resilienza e riguarda la risposta del carico di lavoro in caso di emergenza (un disastro è un evento che causa un grave impatto negativo sull'azienda). Questa risposta deve basarsi sugli obiettivi aziendali dell'organizzazione, che specificano la strategia del carico di lavoro per evitare la perdita di dati, nota come Recovery Point Objective (RPO), e ridurre i tempi di inattività laddove il carico di lavoro non è disponibile per l'uso, nota come Recovery Time Objective (RTO). È quindi necessario implementare la resilienza nella progettazione dei carichi di lavoro nel cloud per raggiungere gli obiettivi di ripristino (RPO e RTO) per un determinato evento di emergenza occasionale. Questo approccio aiuta l'organizzazione a mantenere la continuità aziendale come parte del Business Continuity Planning (BCP).

Questo paper si concentra su come pianificare, progettare e implementare architetture AWS che soddisfino gli obiettivi di disaster recovery per l'azienda. Le informazioni qui condivise sono destinate a coloro che ricoprono ruoli tecnologici, ad esempio Chief Technology Officer (CTOs), architetti, sviluppatori, membri del team operativo e a coloro che hanno il compito di valutare e mitigare i rischi.

Disaster recovery e disponibilità

Il disaster recovery può essere paragonato alla disponibilità, che è un altro componente importante della strategia di resilienza. Mentre il disaster recovery misura gli obiettivi per eventi occasionali, gli obiettivi di disponibilità misurano i valori medi su un periodo di tempo.

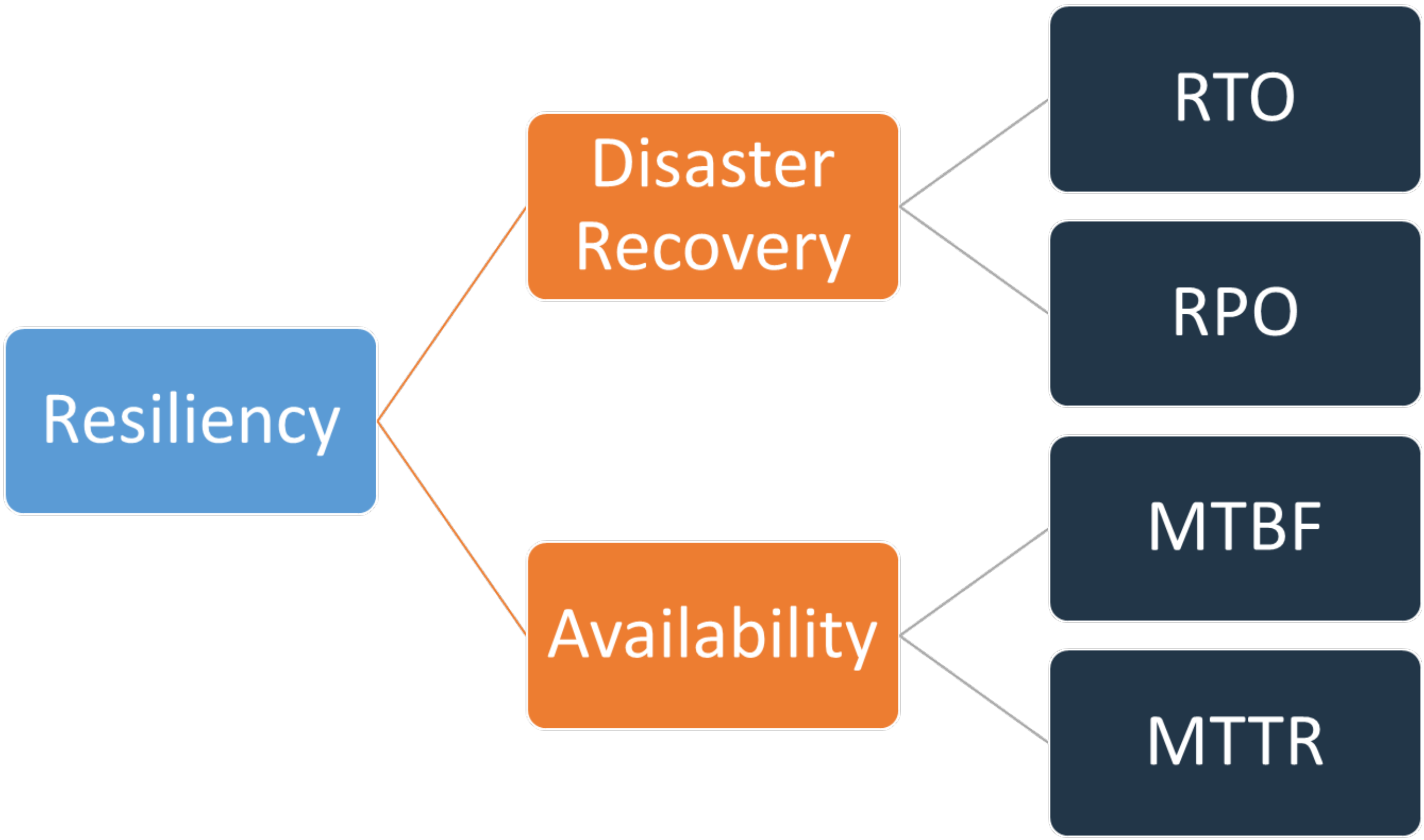

Figura 1 - Obiettivi di resilienza

La disponibilità viene calcolata utilizzando Mean Time Between Failures (MTBF) e Mean Time to Recover (MTTR):

$$Availability = \frac{Available\ for\ Use\ Time}{Total\ Time} = \frac{MTBF}{MTBF + MTTR}$$

Questo approccio viene spesso definito «nove», mentre un obiettivo di disponibilità del 99,9% viene definito «tre nove».

Per quanto riguarda il carico di lavoro, potrebbe essere più semplice contare le richieste riuscite e quelle non riuscite invece di utilizzare un approccio basato sul tempo. In questo caso, è possibile utilizzare il seguente calcolo:

$$Availability = \frac{Successful\ Responses}{Valid\ Requests}$$

Il disaster recovery si concentra sugli eventi di emergenza, mentre la disponibilità si concentra sulle interruzioni più comuni su piccola scala, come guasti dei componenti, problemi di rete, bug software e picchi di carico. L'obiettivo del disaster recovery è la continuità aziendale, mentre la disponibilità riguarda la massimizzazione del tempo di disponibilità di un carico di lavoro per eseguire le funzionalità aziendali previste. Entrambi dovrebbero far parte della vostra strategia di resilienza.

Sei tu Well-Architected?

AWS Well-Architected Framework ti aiuta a comprendere i pro e i contro delle decisioni che prendi quando crei sistemi nel cloud. I sei pilastri del Framework ti consentono di apprendere le migliori pratiche architettoniche per progettare e gestire sistemi affidabili, sicuri, efficienti, convenienti e sostenibili. Utilizzando AWS Well-Architected Tool, disponibile gratuitamente nella Console di gestione AWS, puoi esaminare i tuoi carichi di lavoro rispetto a queste best practice rispondendo a una serie di domande per ogni pilastro.

I concetti trattati in questo white paper ampliano le migliori pratiche contenute nel white paper Reliability Pillar, in particolare la domanda REL 13, «How do you plan for disaster recovery (DR)?». Dopo aver implementato le pratiche in questo white paper, assicurati di rivedere (o rivedere) il tuo carico di lavoro utilizzando AWS Well-Architected Tool.

Modello di responsabilità condivisa per la resilienza

La resilienza è una responsabilità condivisa tra te AWS e il cliente. È importante comprendere in che modo il disaster recovery e la disponibilità, come parte della resilienza, operano nell'ambito di questo modello condiviso.

Responsabilità di AWS «Resilienza del cloud»

AWS è responsabile della resilienza dell'infrastruttura che gestisce tutti i servizi offerti nel cloud AWS. Questa infrastruttura comprende l'hardware, il software, la rete e le strutture che eseguono i servizi cloud AWS. AWS compie sforzi commercialmente ragionevoli per rendere disponibili questi servizi cloud AWS, garantendo che la disponibilità del servizio soddisfi o superi gli AWS Service Level Agreement () SLAs.

L'infrastruttura cloud globale AWS è progettata per consentire ai clienti di creare architetture di carichi di lavoro altamente resilienti. Ogni regione AWS è completamente isolata ed è composta da più zone di disponibilità, che sono partizioni dell'infrastruttura fisicamente isolate. Le zone di disponibilità isolano gli errori che potrebbero influire sulla resilienza del carico di lavoro, impedendo loro di interessare altre zone nella regione. Allo stesso tempo, tutte le zone di una regione AWS sono interconnesse con reti ad alta larghezza di banda e bassa latenza, tramite fibra metropolitana dedicata e completamente ridondante che fornisce reti ad alto throughput e bassa latenza tra le zone. Tutto il traffico tra zone è crittografato. Le prestazioni di rete sono adeguate per l'esecuzione della replica sincrona tra zone. Quando un'applicazione è partizionata AZs, le aziende sono meglio isolate e protette da problemi come interruzioni di corrente, fulmini, tornado, uragani e altro ancora.

Responsabilità del cliente «Resilienza nel cloud»

La tua responsabilità sarà determinata dai servizi cloud AWS selezionati. La scelta definisce l'entità delle attività di configurazione che devi eseguire nell'ambito delle tue responsabilità nell'ambito della resilienza. Ad esempio, un servizio come Amazon Elastic Compute Cloud (Amazon EC2) richiede al cliente di eseguire tutte le attività di configurazione e gestione della resilienza necessarie. I clienti che distribuiscono EC2 istanze Amazon sono responsabili della distribuzione delle EC2 istanze in più sedi (come le zone di disponibilità AWS), dell'implementazione della riparazione automatica utilizzando servizi come Amazon Auto Scaling EC2 e dell'utilizzo delle best practice di architettura resiliente per i carichi di lavoro per le applicazioni installate sulle istanze. Per i servizi gestiti, come Amazon S3 e Amazon DynamoDB, AWS gestisce il livello di infrastruttura, il sistema operativo e le piattaforme e i clienti accedono agli endpoint per archiviare e recuperare i dati. Tu hai la responsabilità della gestione della resilienza dei dati, incluse le strategie di backup, controllo delle versioni e replica.

La distribuzione del carico di lavoro su più zone di disponibilità in una regione AWS fa parte di una strategia di alta disponibilità progettata per proteggere i carichi di lavoro isolando i problemi in una zona di disponibilità e utilizza la ridondanza delle altre zone di disponibilità per continuare a soddisfare le richieste. Un'architettura multi-AZ è parte anche di una strategia di disaster recovery progettata per isolare e proteggere meglio i carichi di lavoro da problemi come le interruzioni dell'alimentazione, i fulmini, i tornado, i terremoti e altri ancora. Le strategie di DR possono anche fare uso di più regioni AWS. Ad esempio, in una configurazione attiva/passiva, il servizio per il carico di lavoro eseguirà il failover dalla regione attiva alla regione DR se la regione attiva non è più in grado di soddisfare le richieste.

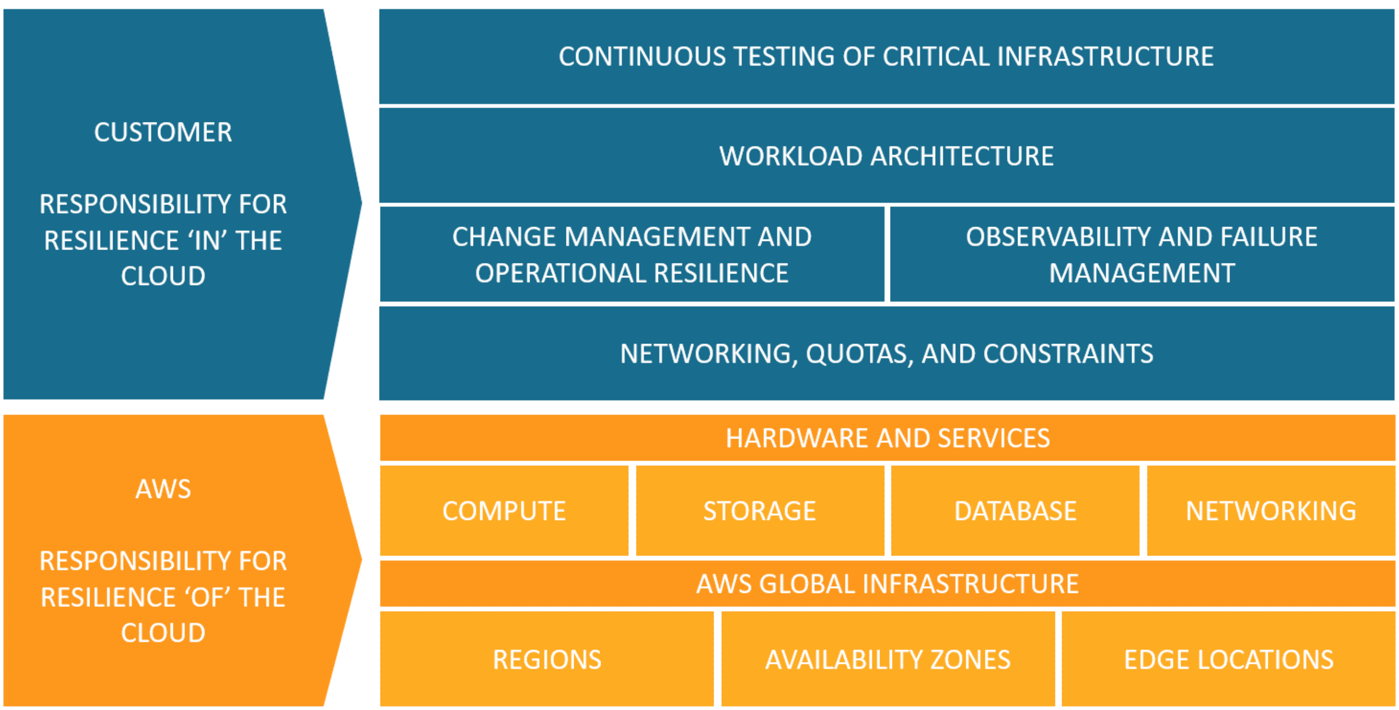

Figura 2 - La resilienza è una responsabilità condivisa tra AWS e il cliente

Cos'è un disastro?

Quando pianifichi il disaster recovery, valuta il tuo piano per queste tre categorie principali di disastri:

- Disastri naturali, come terremoti o inondazioni
- Guasti tecnici, come interruzione dell'alimentazione o della connettività di rete
- Azioni umane, come errori di configurazione involontari o accesso o modifica di parti unauthorized/ outside

Ciascuno di questi potenziali disastri avrà anche un impatto geografico che può essere locale, regionale, nazionale, continentale o globale. Sia la natura del disastro che l'impatto geografico sono importanti quando si considera la strategia di disaster recovery. Ad esempio, è possibile mitigare un problema di inondazione locale che causa un'interruzione del data center utilizzando una strategia Multi-AZ, poiché non influirebbe su più di una zona di disponibilità. Tuttavia, un attacco ai dati di produzione richiederebbe di invocare una strategia di disaster recovery che esegua il failover per eseguire il backup dei dati in un'altra regione AWS.

L'alta disponibilità non è il disaster recovery

Sia la disponibilità che il disaster recovery si basano su alcune delle stesse best practice, come il monitoraggio degli errori, l'implementazione su più sedi e il failover automatico. Tuttavia, Availability si concentra sui componenti del carico di lavoro, mentre il disaster recovery si concentra su copie discrete dell'intero carico di lavoro. Il disaster recovery ha obiettivi diversi dall'Availability, in quanto misura il tempo necessario per il ripristino dopo eventi su larga scala che si qualificano come disastri. È innanzitutto necessario assicurarsi che il carico di lavoro soddisfi gli obiettivi di disponibilità, poiché un'architettura ad alta disponibilità consentirà di soddisfare le esigenze dei clienti in caso di eventi che influiscono sulla disponibilità. La strategia di disaster recovery richiede approcci diversi da quelli per la disponibilità, incentrati sull'implementazione di sistemi discreti in più sedi, in modo da poter eseguire il failover dell'intero carico di lavoro, se necessario.

È necessario considerare la disponibilità del carico di lavoro nella pianificazione del disaster recovery, poiché influirà sull'approccio adottato. Un carico di lavoro eseguito su una singola EC2 istanza Amazon in una zona di disponibilità non ha una disponibilità elevata. Se un problema di allagamento locale riguarda quella zona di disponibilità, questo scenario richiede il failover su un'altra zona di emergenza per soddisfare gli obiettivi di disaster recovery. Confronta questo scenario con un carico di lavoro ad alta disponibilità distribuito attivo/attivo su più siti, in cui il carico di lavoro viene distribuito su più regioni attive e tutte le regioni servono il traffico di produzione. In questo caso, anche nell'improbabile eventualità che un grave disastro renda inutilizzabile una regione, la strategia DR viene realizzata instradando tutto il traffico verso le regioni rimanenti.

Il modo in cui si affrontano i dati è diverso anche tra disponibilità e disaster recovery. Prendi in considerazione una soluzione di storage che si replica continuamente su un altro sito per ottenere un'elevata disponibilità (ad esempio un carico di active/active lavoro multisito). Se uno o più file vengono eliminati o danneggiati sul dispositivo di storage principale, tali modifiche distruttive possono essere replicate sul dispositivo di storage secondario. In questo scenario, nonostante l'elevata disponibilità, la capacità di eseguire il failover in caso di cancellazione o danneggiamento dei dati risulterebbe compromessa. Al contrario, è necessario anche un point-in-time backup come parte di una strategia di disaster recovery.

Piano di continuità operativa (BCP)

Il piano di disaster recovery dovrebbe essere un sottoinsieme del piano di continuità aziendale (BCP) dell'organizzazione, non dovrebbe essere un documento a sé stante. Non ha senso mantenere obiettivi di disaster recovery aggressivi per il ripristino di un carico di lavoro se gli obiettivi aziendali di tale carico di lavoro non possono essere raggiunti a causa dell'impatto del disastro su elementi dell'azienda diversi dal carico di lavoro. Ad esempio, un terremoto potrebbe impedirti di trasportare i prodotti acquistati sulla tua applicazione di eCommerce: anche se un DR efficace mantiene il carico di lavoro funzionante, il tuo BCP deve soddisfare le esigenze di trasporto. La strategia di disaster recovery deve basarsi sui requisiti, sulle priorità e sul contesto aziendali.

Analisi dell'impatto aziendale e valutazione del rischio

Un'analisi dell'impatto aziendale dovrebbe quantificare l'impatto aziendale di un'interruzione dei carichi di lavoro. Dovrebbe identificare l'impatto sui clienti interni ed esterni dell'impossibilità di utilizzare i carichi di lavoro e l'effetto che ciò ha sulla vostra attività. L'analisi dovrebbe aiutare a determinare la rapidità con cui il carico di lavoro deve essere reso disponibile e la quantità di perdita di dati che può essere tollerata. Tuttavia, è importante notare che gli obiettivi di ripristino non devono essere fissati isolatamente; la probabilità di interruzione e il costo del ripristino sono fattori chiave che contribuiscono a determinare il valore aziendale della fornitura di disaster recovery per un carico di lavoro.

L'impatto aziendale può dipendere dal tempo. Potresti prendere in considerazione la possibilità di tenere conto di questo aspetto nella tua pianificazione del disaster recovery. Ad esempio, è probabile che l'interruzione del sistema di gestione delle retribuzioni abbia un impatto molto forte sull'azienda appena prima che tutti vengano pagati, ma può avere un impatto minore subito dopo che tutti sono già stati pagati.

Una valutazione del rischio del tipo di disastro e dell'impatto geografico, insieme a una panoramica dell'implementazione tecnica del carico di lavoro, determinerà la probabilità che si verifichino interruzioni per ogni tipo di emergenza.

Per carichi di lavoro altamente critici, potresti prendere in considerazione l'implementazione dell'infrastruttura in più regioni con replica dei dati e backup continui per ridurre al minimo l'impatto aziendale. Per i carichi di lavoro meno critici, una strategia valida potrebbe essere quella di non implementare affatto il disaster recovery. Inoltre, per alcuni scenari di emergenza, è utile non adottare alcuna strategia di disaster recovery che consenta di prendere una decisione informata basata su una bassa probabilità che il disastro si verifichi. Ricorda che le zone di disponibilità all'interno di una regione AWS sono già progettate con una distanza significativa tra loro e un'attenta pianificazione della posizione, in modo che i disastri più comuni abbiano un impatto solo su una zona e non sulle altre. Pertanto, un'architettura Multi-AZ all'interno di una regione AWS potrebbe già soddisfare gran parte delle tue esigenze di mitigazione del rischio.

Il costo delle opzioni di disaster recovery deve essere valutato per garantire che la strategia di disaster recovery fornisca il giusto livello di valore aziendale considerando l'impatto e il rischio aziendali.

Con tutte queste informazioni, è possibile documentare la minaccia, il rischio, l'impatto e il costo dei diversi scenari di emergenza e le opzioni di ripristino associate. Queste informazioni devono essere utilizzate per determinare gli obiettivi di ripristino per ciascuno dei carichi di lavoro.

Obiettivi di ripristino (RTO e RPO)

Quando si crea una strategia di Disaster Recovery (DR), le organizzazioni generalmente pianificano il Recovery Time Objective (RTO) e il Recovery Point Objective (RPO).

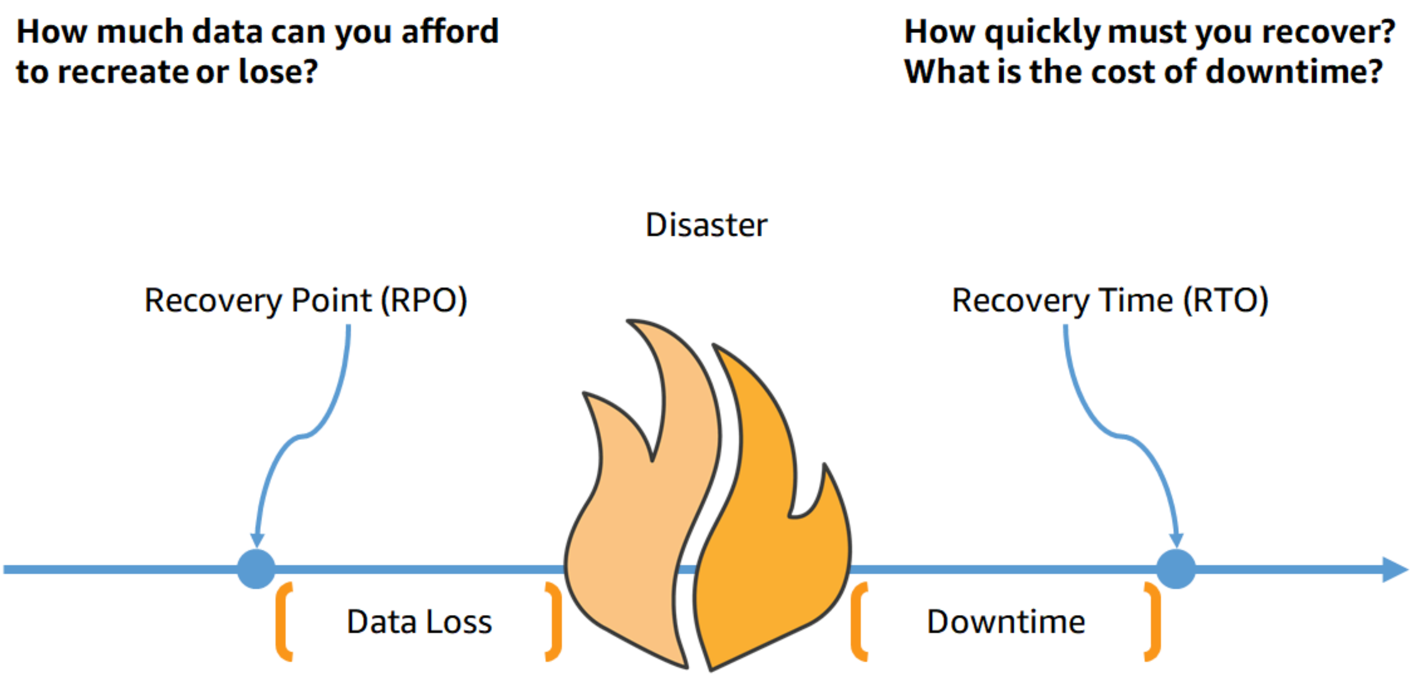

Figura 3 - Obiettivi di ripristino

Il Recovery Time Objective (RTO) è il ritardo massimo accettabile tra l'interruzione del servizio e il ripristino del servizio. Questo obiettivo determina quale finestra temporale è considerata accettabile quando il servizio non è disponibile ed è definito dall'organizzazione.

In questo paper vengono discusse principalmente quattro strategie di disaster recovery: backup e ripristino, pilot light, warm standby e multisito active/active (vedi [Opzioni di disaster recovery nel cloud](#)). Nel diagramma seguente, l'azienda ha determinato l'RTO massimo consentito e il limite di quanto può spendere per la strategia di ripristino dei servizi. Considerati gli obiettivi aziendali, le strategie DR Pilot Light o Warm Standby soddisferanno sia l'RTO che i criteri di costo.

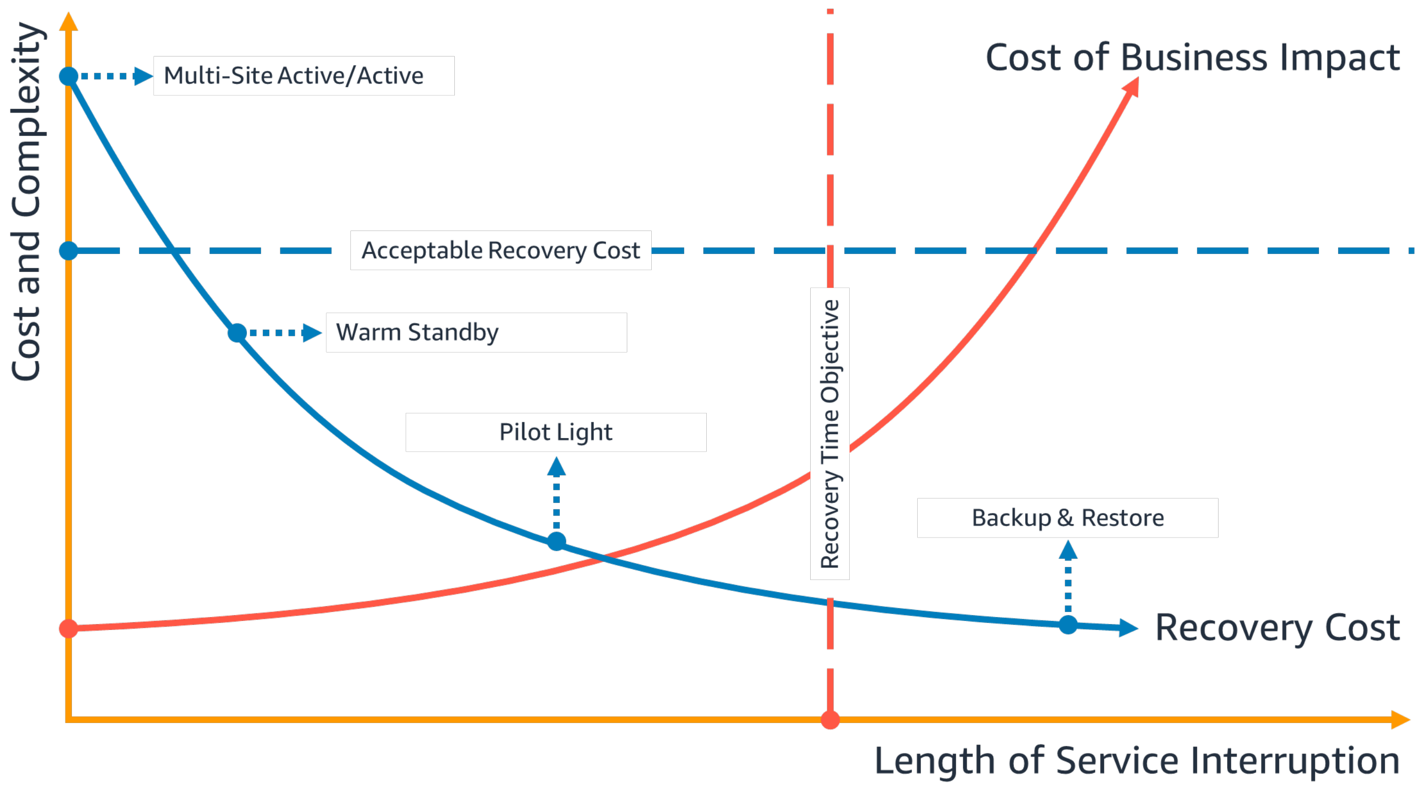

Figura 4 - Obiettivo del tempo di ripristino

Il Recovery Point Objective (RPO) è il periodo di tempo massimo accettabile dall'ultimo punto di ripristino dei dati. Questo obiettivo determina ciò che è considerato una perdita accettabile di dati tra l'ultimo punto di ripristino e l'interruzione del servizio ed è definito dall'organizzazione.

Nel diagramma seguente, l'azienda ha determinato l'RPO massimo consentito e il limite di quanto può spendere per la propria strategia di ripristino dei dati. Delle quattro strategie DR, Pilot Light o Warm Standby DR soddisfano entrambi i criteri di RPO e di costo.

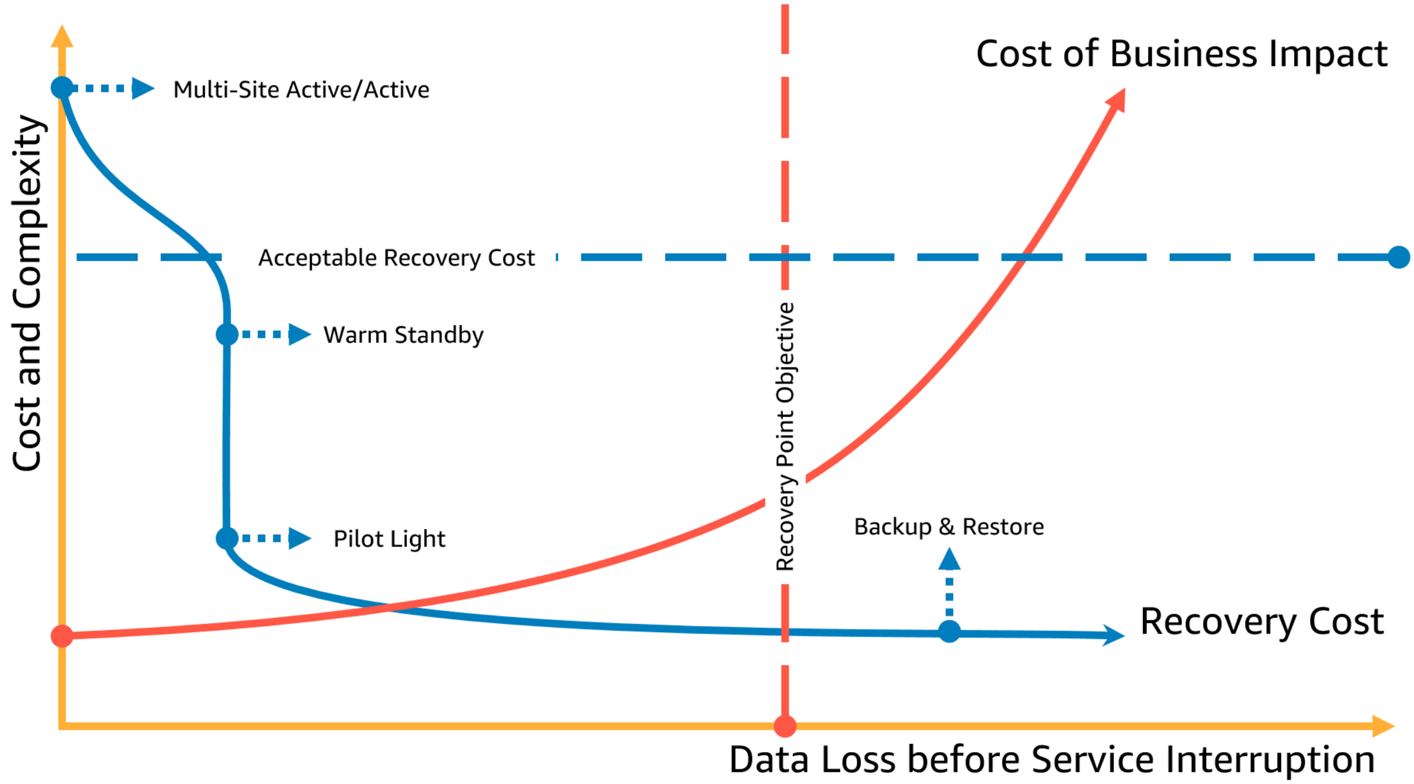

Figura 5 - Obiettivo del punto di ripristino

> **Note**
>
> Se il costo della strategia di ripristino è superiore al costo dell'errore o della perdita, l'opzione di ripristino non dovrebbe essere adottata a meno che non vi sia un fattore secondario, come i requisiti normativi. Quando effettui questa valutazione, prendi in considerazione strategie di ripristino a costi variabili.

Il disaster recovery è diverso nel cloud

Le strategie di disaster recovery si evolvono con l'innovazione tecnica. Un piano di disaster recovery locale può comportare il trasporto fisico dei nastri o la replica dei dati su un altro sito. La tua organizzazione deve rivalutare l'impatto aziendale, il rischio e il costo delle sue precedenti strategie di disaster recovery per raggiungere i suoi obiettivi di DR su AWS. Il disaster recovery nel cloud AWS include i seguenti vantaggi rispetto agli ambienti tradizionali:

- Recupera rapidamente da un disastro con una complessità ridotta
- I test semplici e ripetibili consentono di eseguire i test più facilmente e con maggiore frequenza
- Un sovraccarico di gestione inferiore riduce il carico operativo
- Le opportunità di automazione riducono le possibilità di errore e migliorano i tempi di ripristino

AWS consente di scambiare le spese fisse di capitale di un data center di backup fisico con le spese operative variabili di un ambiente cloud di dimensioni adeguate, il che può ridurre significativamente i costi.

Per molte organizzazioni, il disaster recovery locale si basava sul rischio di interruzione di uno o più carichi di lavoro in un data center e sul ripristino dei dati di backup o replicati in un data center secondario. Quando le organizzazioni distribuiscono carichi di lavoro su AWS, possono implementare un carico di lavoro ben architettato e fare affidamento sul design dell'infrastruttura cloud globale di AWS per mitigare l'effetto di tali interruzioni. Consulta il white paper AWS Well-Architected Framework - Reliability Pillar per ulteriori informazioni sulle best practice architettoniche per progettare e gestire carichi di lavoro affidabili, sicuri, efficienti ed economici nel cloud. Utilizzalo AWS Well-Architected Toolper rivedere periodicamente i tuoi carichi di lavoro per assicurarti che seguano le migliori pratiche e le linee guida del Well-Architected Framework. Lo strumento è disponibile gratuitamente in. Console di gestione AWS

Se i tuoi carichi di lavoro sono su AWS, non devi preoccuparti della connettività del data center (ad eccezione della possibilità di accedervi), dell'alimentazione, dell'aria condizionata, della soppressione degli incendi e dell'hardware. Tutto questo è gestito per te e hai accesso a più zone di disponibilità isolate dai guasti (ognuna composta da uno o più data center discreti).

Singola regione AWS

Per un evento di emergenza basato sull'interruzione o la perdita di un data center fisico, l'implementazione di un carico di lavoro ad alta disponibilità in più zone di disponibilità all'interno di una singola regione AWS aiuta a mitigare i disastri naturali e tecnici. Il backup continuo dei dati all'interno di questa singola regione può ridurre il rischio di minacce umane, come errori o attività non autorizzate che potrebbero causare la perdita di dati. Ogni regione AWS è composta da più zone di disponibilità, ciascuna isolata dai guasti nelle altre zone. Ogni zona di disponibilità è a sua volta composta da uno o più data center fisici discreti. Per isolare meglio i problemi più importanti e ottenere un'elevata disponibilità, puoi partizionare i carichi di lavoro su più zone della stessa regione. Le zone di disponibilità sono progettate per la ridondanza fisica e forniscono resilienza, consentendo prestazioni ininterrotte, anche in caso di interruzioni di corrente, interruzioni di Internet, inondazioni e altri disastri naturali. Consulta AWS Global Cloud Infrastructure per scoprire come AWS lo fa.

Implementando su più zone di disponibilità in una singola regione AWS, il carico di lavoro è protetto meglio dai guasti di un singolo (o anche più) data center. Per una maggiore sicurezza con la distribuzione in un'unica regione, puoi eseguire il backup dei dati e della configurazione (inclusa la definizione dell'infrastruttura) in un'altra regione. Questa strategia riduce l'ambito del piano di disaster recovery e include solo il backup e il ripristino dei dati. Sfruttare la resilienza multiregionale eseguendo il backup su un'altra regione AWS è semplice ed economico rispetto alle altre opzioni multiregionali descritte nella sezione seguente. Ad esempio, il backup su Amazon Simple Storage Service (Amazon S3) ti consente di accedere al recupero immediato dei tuoi dati. Tuttavia, se la tua strategia di ripristino di emergenza per porzioni dei dati prevede requisiti più ridotti per i tempi di recupero (da minuti a ore), l'utilizzo di Amazon Glacier o Amazon Glacier Deep Archive ridurrà in modo significativo i costi della tua strategia di backup e ripristino.

Alcuni carichi di lavoro possono avere requisiti normativi di residenza dei dati. Se ciò si applica al tuo carico di lavoro in una località che attualmente ha una sola regione AWS, oltre a progettare carichi di lavoro Multi-AZ per l'alta disponibilità come discusso sopra, puoi anche utilizzare i carichi di lavoro AZs all'interno di quella regione come postazioni discrete, il che può essere utile per soddisfare i requisiti di residenza dei dati applicabili al tuo carico di lavoro all'interno di quella regione. Le strategie di DR descritte nelle sezioni seguenti utilizzano più regioni AWS, ma possono anche essere implementate utilizzando zone di disponibilità anziché regioni.

Più regioni AWS

Per un evento di emergenza che include il rischio di perdere più data center a una distanza significativa l'uno dall'altro, dovresti prendere in considerazione opzioni di disaster recovery per mitigare i disastri naturali e tecnici che colpiscono un'intera regione all'interno di AWS. Tutte le opzioni descritte nelle seguenti sezioni possono essere implementate come architetture multiregionali per proteggersi da tali disastri.

Opzioni di disaster recovery nel cloud

Le strategie di disaster recovery disponibili in AWS possono essere ampiamente suddivise in quattro approcci, che vanno dal basso costo e dalla bassa complessità dei backup a strategie più complesse che utilizzano più regioni attive. Active/passive le strategie utilizzano un sito attivo (come una regione AWS) per ospitare il carico di lavoro e servire il traffico. Il sito passivo (ad esempio un'altra regione AWS) viene utilizzato per il ripristino. Il sito passivo non serve attivamente il traffico fino a quando non viene attivato un evento di failover.

È fondamentale valutare e testare regolarmente la strategia di disaster recovery in modo da poterla utilizzare con fiducia, se necessario. Usa [AWS Resilience Hub] per convalidare e monitorare continuamente la resilienza dei tuoi AWS carichi di lavoro, inclusa la probabilità di raggiungere gli obiettivi RTO e RPO.

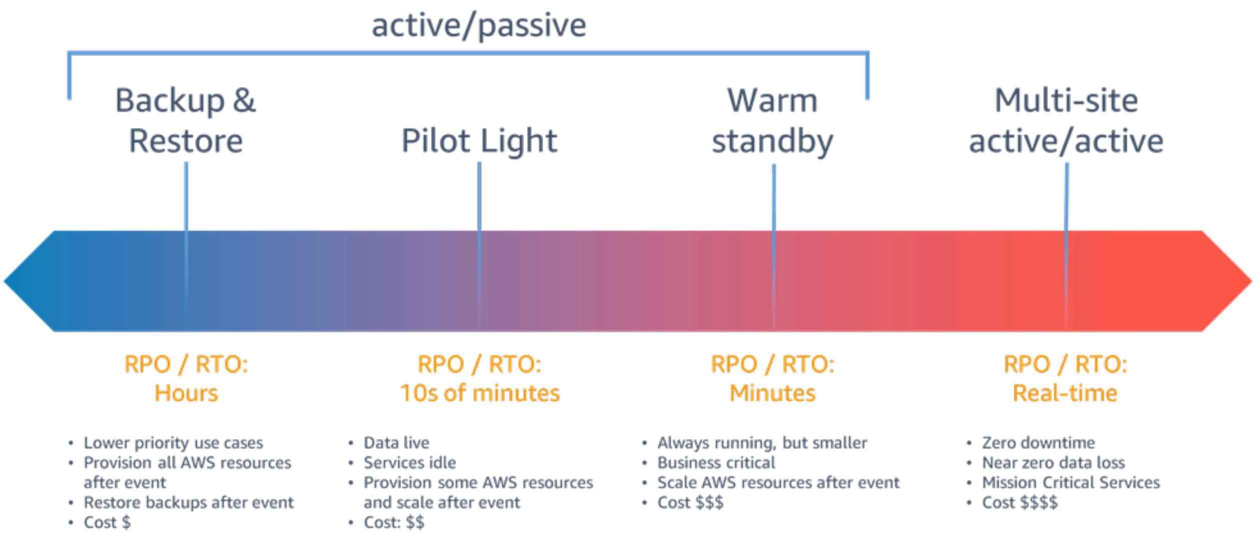

Figura 6 - Strategie di disaster recovery

Per un evento di emergenza basato sull'interruzione o la perdita di un data center fisico per un carico di lavoro [ben progettato] e ad alta disponibilità, può essere necessario solo un approccio di backup e ripristino al disaster recovery. Se la definizione di disastro non si limita all'interruzione o alla perdita di un data center fisico ma si limita a quella di una regione o se si è soggetti a requisiti normativi che lo richiedono, è consigliabile prendere in considerazione Pilot Light, Warm Standby o Multi-Site Active/Active.

Quando scegli la tua strategia e le risorse AWS per implementarla, tieni presente che all'interno di AWS, di solito dividiamo i servizi in piano dati e piano di controllo. Il piano dati è responsabile della fornitura del servizio in tempo reale mentre i piani di controllo (control-plane) consentono di configurare l'ambiente. Per la massima resilienza, è necessario utilizzare solo le operazioni del piano dati come parte delle operazioni di failover. Questo perché i piani dati in genere hanno obiettivi di progettazione di disponibilità più elevati rispetto ai piani di controllo.

Backup e ripristino

Il backup e il ripristino sono un approccio adatto per mitigare la perdita o il danneggiamento dei dati. Questo approccio può essere utilizzato anche per mitigare un disastro regionale replicando i dati in altre regioni AWS o per mitigare la mancanza di ridondanza per i carichi di lavoro distribuiti in una singola zona di disponibilità. Oltre ai dati, è necessario ridistribuire l'infrastruttura, la configurazione e il codice dell'applicazione nella regione di ripristino. Per consentire una ridistribuzione rapida dell'infrastruttura senza errori, è consigliabile eseguire sempre la distribuzione utilizzando Infrastructure as Code (IaC) utilizzando servizi come o il. AWS CloudFormationAWS Cloud Development Kit (AWS CDK) Senza IaC, potrebbe essere complesso ripristinare i carichi di lavoro nella regione di ripristino, il che comporterà un aumento dei tempi di ripristino e probabilmente il superamento dell'RTO. Oltre ai dati utente, assicurati di eseguire anche il backup del codice e della configurazione, tra cui Amazon Machine Images (AMIs) che usi per creare EC2 istanze Amazon. Puoi utilizzarlo AWS CodePipelineper automatizzare la ridistribuzione del codice e della configurazione dell'applicazione.

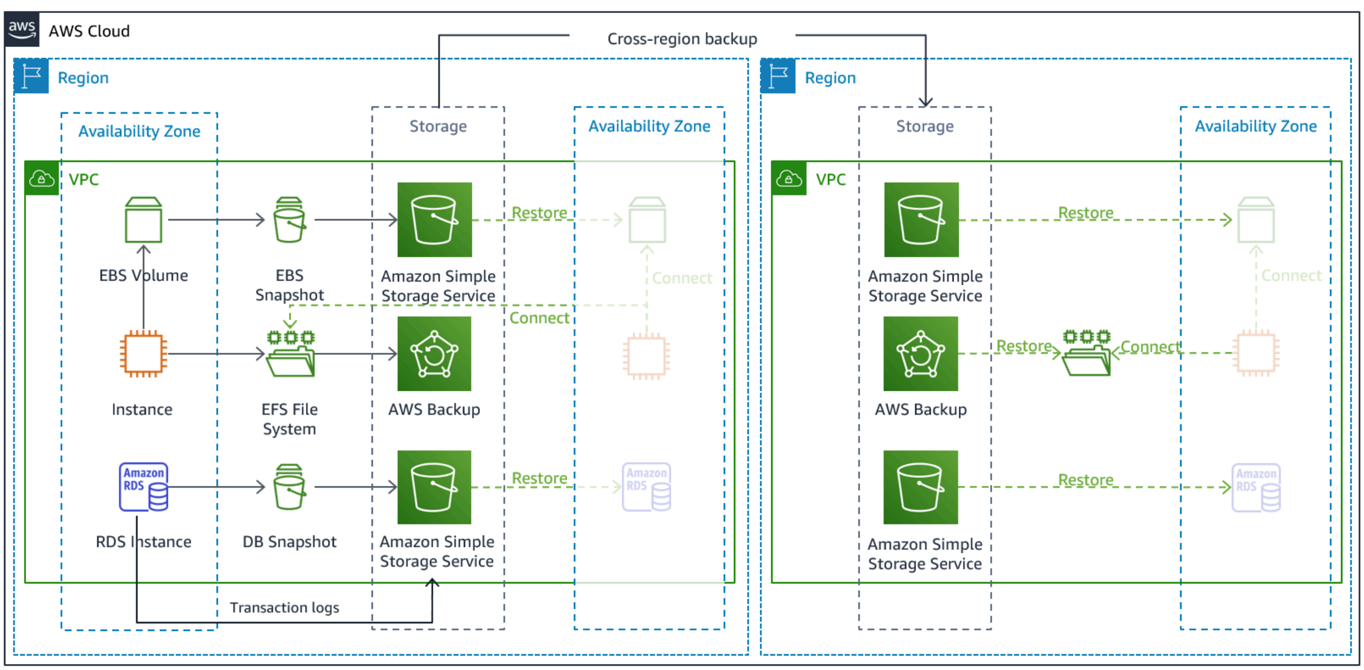

Figura 7 - Architettura di backup e ripristino

Servizi AWS

I dati del carico di lavoro richiederanno una strategia di backup che venga eseguita periodicamente o sia continua. La frequenza con cui esegui il backup determinerà il punto di ripristino raggiungibile (che deve essere allineato per soddisfare l'RPO). Il backup dovrebbe inoltre offrire un modo per ripristinarlo al momento in cui è stato eseguito. Il backup con point-in-time ripristino è disponibile tramite i seguenti servizi e risorse:

- Istantanea di Amazon Elastic Block Store (Amazon EBS)
- Backup con Amazon DynamoDB
- Istantanea Amazon RDS
- Istantanea di Amazon Aurora DB
- Backup Amazon EFS (se utilizzato AWS Backup)
- Istantanea di Amazon Redshift
- Istantanea di Amazon Neptune
- Amazon DocumentDB
- Amazon FSx per Windows File Server, Amazon FSx for Lustre, Amazon FSx per NetApp ONTAP e Amazon FSx per OpenZFS

Per Amazon Simple Storage Service (Amazon S3) Simple Storage Service (Amazon S3), puoi utilizzare Amazon S3 Cross-Region Replication (CRR) per copiare in modo asincrono oggetti su un bucket S3 nella regione DR in modo continuo, fornendo al contempo il controllo delle versioni per gli oggetti archiviati in modo da poter scegliere il punto di ripristino. La replica continua dei dati ha il vantaggio di essere il tempo più breve (quasi zero) per eseguire il backup dei dati, ma potrebbe non proteggere da eventi di emergenza come il danneggiamento dei dati o attacchi dolosi (come l'eliminazione non autorizzata dei dati) e dai point-in-time backup. La replica continua è trattata nella sezione AWS Services for Pilot Light.

AWS Backupfornisce una posizione centralizzata per configurare, pianificare e monitorare le funzionalità di backup di AWS per i seguenti servizi e risorse:

- Volumi Amazon Elastic Block Store (Amazon EBS)
- EC2Istanze Amazon
- Database Amazon Relational Database Service (Amazon RDS) (inclusi i database Amazon Aurora)

- Tabelle Amazon DynamoDB
- File system Amazon Elastic File System (Amazon EFS)
- Volumi Gateway di archiviazione AWS
- Amazon FSx per Windows File Server, Amazon FSx for Lustre, Amazon FSx per NetApp ONTAP e Amazon FSx per OpenZFS

AWS Backup supporta la copia di backup tra regioni, ad esempio in una regione di disaster recovery.

Come strategia di disaster recovery aggiuntiva per i tuoi dati Amazon S3, abilita il controllo delle versioni degli oggetti S3. Il controllo delle versioni degli oggetti protegge i dati in S3 dalle conseguenze delle azioni di eliminazione o modifica conservando la versione originale prima dell'azione. Il controllo delle versioni degli oggetti può essere un'utile mitigazione per i disastri di tipo umano. Se utilizzi la replica S3 per eseguire il backup dei dati nella tua regione DR, per impostazione predefinita, quando un oggetto viene eliminato nel bucket di origine, Amazon S3 aggiunge un marker di eliminazione solo nel bucket di origine. Questo approccio protegge i dati nella regione DR da eliminazioni dannose nella regione di origine.

Oltre ai dati, è necessario eseguire il backup della configurazione e dell'infrastruttura necessarie per ridistribuire il carico di lavoro e raggiungere il Recovery Time Objective (RTO). AWS CloudFormationfornisce Infrastructure as Code (IaC) e ti consente di definire tutte le risorse AWS nel tuo carico di lavoro in modo da poterlo distribuire e ridistribuire in modo affidabile su più account AWS e regioni AWS. Puoi eseguire il backup EC2 delle istanze Amazon utilizzate dal tuo carico di lavoro come Amazon Machine Images ()AMIs. L'AMI viene creata da istantanee del volume root dell'istanza e di qualsiasi altro volume EBS collegato all'istanza. Puoi utilizzare questa AMI per avviare una versione ripristinata dell' EC2 istanza. Un AMI può essere copiato all'interno o tra le regioni. In alternativa, puoi utilizzare AWS Backupper copiare i backup tra account e in altre regioni AWS. La funzionalità di backup su più account aiuta a proteggere da eventi di emergenza che includono minacce interne o compromissione dell'account. AWS Backup aggiunge inoltre funzionalità aggiuntive per il EC2 backup: oltre ai singoli volumi EBS dell'istanza, AWS Backup archivia e tiene traccia dei seguenti metadati: tipo di istanza, cloud privato virtuale (VPC) configurato, gruppo di sicurezza, ruolo IAM, configurazione di monitoraggio e tag. Tuttavia, questi metadati aggiuntivi vengono utilizzati solo per il ripristino del EC2 backup nella stessa regione AWS.

Tutti i dati archiviati nella regione di disaster recovery come backup devono essere ripristinati al momento del failover. AWS Backup offre funzionalità di ripristino, ma attualmente non abilita il ripristino programmato o automatico. Puoi implementare il ripristino automatico nella regione DR utilizzando l'SDK AWS da APIs richiedere AWS Backup. Puoi impostarlo come un processo ricorrente regolarmente o attivare il ripristino ogni volta che viene completato un backup. La figura seguente mostra un esempio di ripristino automatico utilizzando [Amazon Simple Notification Service (Amazon AWS LambdaSNS)] e. L'implementazione di un ripristino periodico dei dati pianificato è una buona idea in quanto il ripristino dei dati dal backup è un'operazione del piano di controllo. Se questa operazione non fosse disponibile durante un disastro, avresti comunque a disposizione archivi dati utilizzabili creati da un backup recente.

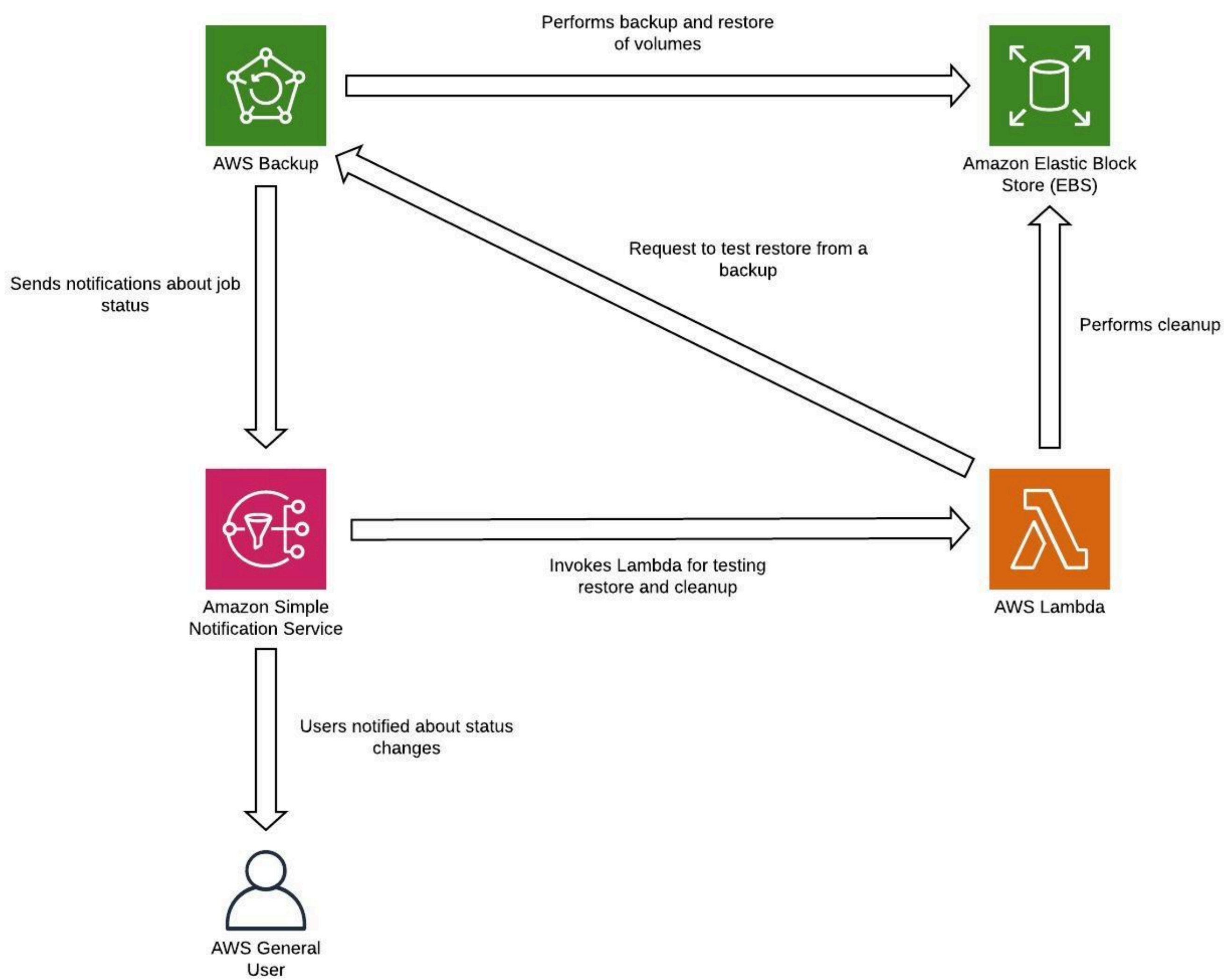

Figura 8 - Ripristino e test dei backup

> **Note**
>
> La tua strategia di backup deve includere il test dei backup. Per ulteriori informazioni, consulta la sezione Testing Disaster Recovery. Fai riferimento a AWS Well-Architected Lab: Testing Backup and Restore of Data per una dimostrazione pratica dell'implementazione.

Pilot light

Con l'approccio pilot light, puoi replicare i dati da una regione all'altra e fornire una copia dell'infrastruttura di carico di lavoro principale. Le risorse necessarie per supportare la replica dei dati e il backup, come database e archiviazione di oggetti, sono sempre attive. Altri elementi, come i server delle applicazioni, vengono caricati con il codice e le configurazioni dell'applicazione, ma sono «disattivati» e vengono utilizzati solo durante i test o quando viene richiamato il failover del disaster recovery. Nel cloud, hai la flessibilità di eseguire il deprovisioning delle risorse quando non ne hai bisogno e il provisioning quando ne hai bisogno. Una procedura consigliata per «disattivarla» consiste nel non distribuire la risorsa e quindi creare la configurazione e le funzionalità necessarie per distribuirla («accenderla») quando necessario. A differenza dell'approccio di backup e ripristino, l'infrastruttura principale è sempre disponibile e si ha sempre la possibilità di fornire rapidamente un ambiente di produzione su vasta scala attivando e scalando i server delle applicazioni.

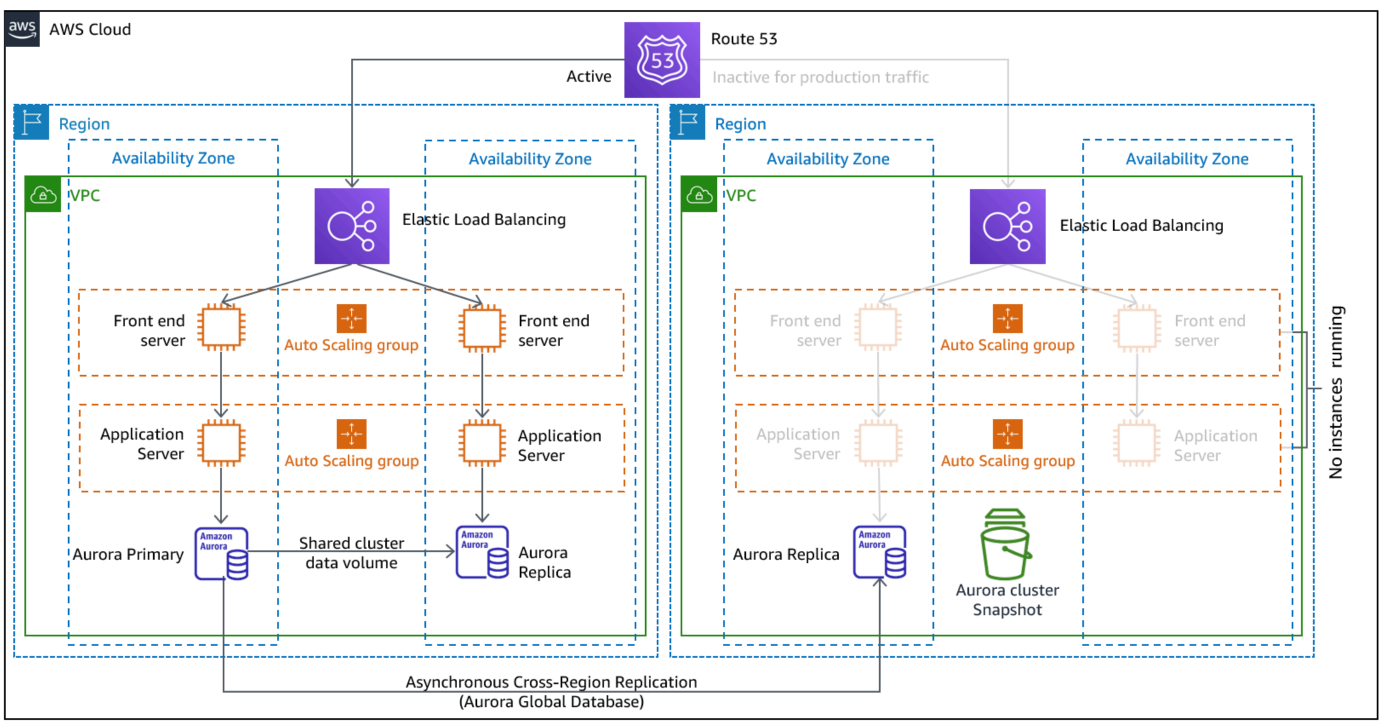

Figura 9 - Architettura dell'illuminazione pilota

Un approccio pilota leggero riduce al minimo i costi correnti del disaster recovery riducendo al minimo le risorse attive e semplifica il ripristino al momento del disastro perché i requisiti dell'infrastruttura di base sono tutti soddisfatti. Questa opzione di ripristino richiede la modifica dell'approccio di implementazione. È necessario apportare modifiche all'infrastruttura di base in ciascuna regione e implementare le modifiche del carico di lavoro (configurazione, codice) contemporaneamente in ciascuna regione. Questo passaggio può essere semplificato automatizzando le implementazioni e utilizzando l'infrastruttura come codice (IaC) per distribuire l'infrastruttura su più account e regioni (implementazione completa dell'infrastruttura nella regione principale e implementazione dell'infrastruttura ridimensionata/disattivata nelle regioni DR). Si consiglia di utilizzare un account diverso per regione per fornire il massimo livello di isolamento delle risorse e della sicurezza (nel caso in cui anche le credenziali compromesse rientrino nei piani di disaster recovery).

Con questo approccio, è inoltre necessario mitigare un problema di dati. La replica continua dei dati ti protegge da alcuni tipi di emergenza, ma potrebbe non proteggerti dal danneggiamento o dalla distruzione dei dati, a meno che la tua strategia non includa anche il controllo delle versioni dei dati archiviati o opzioni di ripristino. point-in-time È possibile eseguire il backup dei dati replicati nella regione di emergenza per creare point-in-time backup nella stessa regione.

Servizi AWS

Oltre a utilizzare i servizi AWS descritti nella sezione Backup and Restore per creare point-in-time backup, prendi in considerazione anche i seguenti servizi per la tua strategia pilota.

Per una fase pilota, la replica continua dei dati su database e archivi di dati attivi nella regione DR è l'approccio migliore per un RPO basso (se utilizzato in aggiunta ai point-in-time backup discussi in precedenza). AWS fornisce una replica dei dati continua, interregionale e asincrona utilizzando i seguenti servizi e risorse:

- Replica di Amazon Simple Storage Service (Amazon S3)
- Amazon RDS legge le repliche
- Database globali Amazon Aurora
- Tabelle globali Amazon DynamoDB
- Cluster globali Amazon DocumentDB
- Datastore globale per Amazon ElastiCache (Redis OSS)

Con la replica continua, le versioni dei dati sono disponibili quasi immediatamente nella regione DR. I tempi di replica effettivi possono essere monitorati utilizzando funzionalità di servizio come S3 Replication Time Control (S3 RTC) per oggetti S3 e funzionalità di gestione dei database globali Amazon Aurora.

Quando si esegue il failover per eseguire il read/write carico di lavoro dalla regione di disaster recovery, è necessario promuovere una replica di lettura RDS per farla diventare l'istanza principale. Per le istanze DB diverse da Aurora, il completamento del processo richiede alcuni minuti e il riavvio fa parte del processo. Per la replica tra regioni (CRR) e il failover con RDS, l'utilizzo del database globale Amazon Aurora offre diversi vantaggi. Il database globale utilizza un'infrastruttura dedicata che lascia i database completamente disponibili per servire la tua applicazione e può replicarsi nella regione secondaria con una latenza tipica inferiore a un secondo (e all'interno di una regione AWS è molto inferiore a 100 millisecondi). Con il database globale di Amazon Aurora, se la tua regione principale subisce un peggioramento delle prestazioni o un'interruzione delle prestazioni, puoi promuovere una delle regioni secondarie affinché si assuma responsabilità di lettura/scrittura in meno di un minuto, anche in caso di interruzione totale a livello regionale. Puoi anche configurare Aurora per monitorare il tempo di ritardo dell'RPO di tutti i cluster secondari per assicurarti che almeno un cluster secondario rimanga all'interno della finestra RPO di destinazione.

È necessario implementare una versione ridotta dell'infrastruttura di carico di lavoro principale con un numero inferiore o inferiore di risorse nella regione DR. Utilizzando AWS CloudFormation, puoi definire la tua infrastruttura e distribuirla in modo coerente tra gli account AWS e tra le regioni AWS. AWS CloudFormation utilizza pseudo parametri predefiniti per identificare l'account AWS e la regione AWS in cui viene distribuito. Pertanto, puoi implementare la logica delle condizioni nei tuoi CloudFormation modelli per distribuire solo la versione ridotta dell'infrastruttura nella regione DR. Ad EC2 esempio, le implementazioni, un'Amazon Machine Image (AMI) fornisce informazioni come la configurazione hardware e il software installato. È possibile implementare una pipeline Image Builder che crei ciò di cui AMIs si ha bisogno e copiarla sia nella regione principale che in quella di backup. Questo aiuta a garantire che queste Golden AMIs abbiano tutto il necessario per ridistribuire o scalare il carico di lavoro in una nuova regione, in caso di emergenza. Le EC2 istanze Amazon vengono distribuite in una configurazione ridotta (meno istanze rispetto alla regione principale). Per scalare l'infrastruttura in modo da supportare il traffico di produzione, consulta Amazon Auto EC2 Scaling nella sezione Warm Standby.

Per una active/passive configurazione come Pilot Light, tutto il traffico viene inizialmente indirizzato alla regione principale e passa alla regione di disaster recovery se la regione principale non è più disponibile. Questa operazione di failover può essere avviata automaticamente o manualmente. Il failover avviato automaticamente sulla base di controlli o allarmi di stato deve essere usato con cautela. Anche utilizzando le migliori pratiche illustrate qui, i tempi di ripristino e il punto di ripristino saranno superiori a zero, con una certa perdita di disponibilità e di dati. Se si esegue il failover quando non è necessario (falso allarme), si subiscono tali perdite. Pertanto si usa spesso il failover avviato manualmente. In questo caso, devi comunque automatizzare i passaggi del failover, in modo che l'avvio manuale si limiti al clic su un pulsante.

Esistono diverse opzioni di gestione del traffico da considerare quando si utilizzano AWS i servizi.

Tra le opzioni, vi è l'utilizzo di Amazon Route 53. Utilizzando Amazon Route 53, puoi associare più endpoint IP in una o più regioni AWS a un nome di dominio Route 53. Quindi, puoi indirizzare il traffico verso l'endpoint appropriato con quel nome di dominio. In caso di failover, è necessario spostare il traffico verso l'endpoint di ripristino e allontanarlo dall'endpoint principale. I controlli di integrità di Amazon Route 53 monitorano questi endpoint. Utilizzando questi controlli di integrità, puoi configurare il failover DNS avviato automaticamente per garantire che il traffico venga inviato solo a endpoint integri, un'operazione altamente affidabile eseguita sul piano dati. Per implementarlo utilizzando il failover avviato manualmente, puoi utilizzare Amazon Application Recovery Controller (ARC). Con ARC, puoi creare controlli di integrità di Route 53 che in realtà non controllano lo stato, ma agiscono invece come interruttori di accensione/spegnimento su cui hai il pieno controllo. Utilizzando l'AWS CLI o l'SDK AWS, puoi eseguire il failover degli script utilizzando questa API del piano dati ad alta disponibilità. Lo script attiva questi switch (i controlli di integrità della Route 53) dicendo a Route 53 di inviare il traffico alla regione di ripristino anziché alla regione principale. Un'altra opzione per il failover avviato manualmente, che alcuni hanno utilizzato, consiste nell'utilizzare una politica di routing ponderata e modificare il peso delle regioni primarie e di ripristino in modo che tutto il traffico venga indirizzato alla regione di ripristino. Tuttavia, tieni presente che si tratta di un'operazione sul piano di controllo e quindi non così resiliente come l'approccio al piano dati che utilizza Amazon Application Recovery Controller (ARC).

Un'altra opzione è quella di utilizzare AWS Global Accelerator. Utilizzando l' AnyCast IP, puoi associare più endpoint in una o più regioni AWS allo stesso indirizzo o agli stessi indirizzi IP pubblici statici. AWS Global Accelerator quindi indirizza il traffico verso l'endpoint appropriato associato a quell'indirizzo. I controlli dello stato di Global Accelerator monitorano gli endpoint. Utilizzando questi controlli di integrità, AWS Global Accelerator verifica lo stato delle applicazioni e indirizza automaticamente il traffico degli utenti verso l'endpoint dell'applicazione integro. Per il failover avviato manualmente, è possibile regolare l'endpoint che riceve il traffico utilizzando le ghiere di controllo, ma tenete presente che si tratta di un'operazione del piano di controllo. Global Accelerator offre latenze inferiori all'endpoint dell'applicazione poiché utilizza l'estesa rete edge di AWS per indirizzare il traffico sulla dorsale della rete AWS il prima possibile. Global Accelerator evita inoltre i problemi di memorizzazione nella cache che possono verificarsi con i sistemi DNS (come Route 53).

Amazon CloudFront offre il failover di origine, in base al quale, se una determinata richiesta all'endpoint primario fallisce, CloudFront indirizza la richiesta all'endpoint secondario. A differenza delle operazioni di failover descritte in precedenza, tutte le richieste successive vengono comunque inviate all'endpoint primario e il failover viene eseguito per ogni richiesta.

AWS Disaster Recovery elastico

AWS Elastic Disaster Recovery (DRS) replica continuamente le applicazioni e i database ospitati sul server da qualsiasi origine AWS utilizzando la replica a livello di blocco del server sottostante. Elastic Disaster Recovery consente di utilizzare una regione Cloud AWS come obiettivo di disaster recovery per un carico di lavoro ospitato in locale o su un altro provider di servizi cloud e il relativo ambiente. Può essere utilizzato anche per il disaster recovery dei carichi di lavoro AWS ospitati se sono costituiti solo da applicazioni e database ospitati su EC2 (ovvero non su RDS). Elastic Disaster Recovery utilizza la strategia Pilot Light, mantenendo una copia dei dati e delle risorse «disattivate» in un Amazon Virtual Private Cloud (Amazon VPC) utilizzato come area di staging. Quando viene attivato un evento di failover, le risorse in fasi vengono utilizzate per creare automaticamente una distribuzione a piena capacità nell'Amazon VPC di destinazione utilizzato come posizione di ripristino.

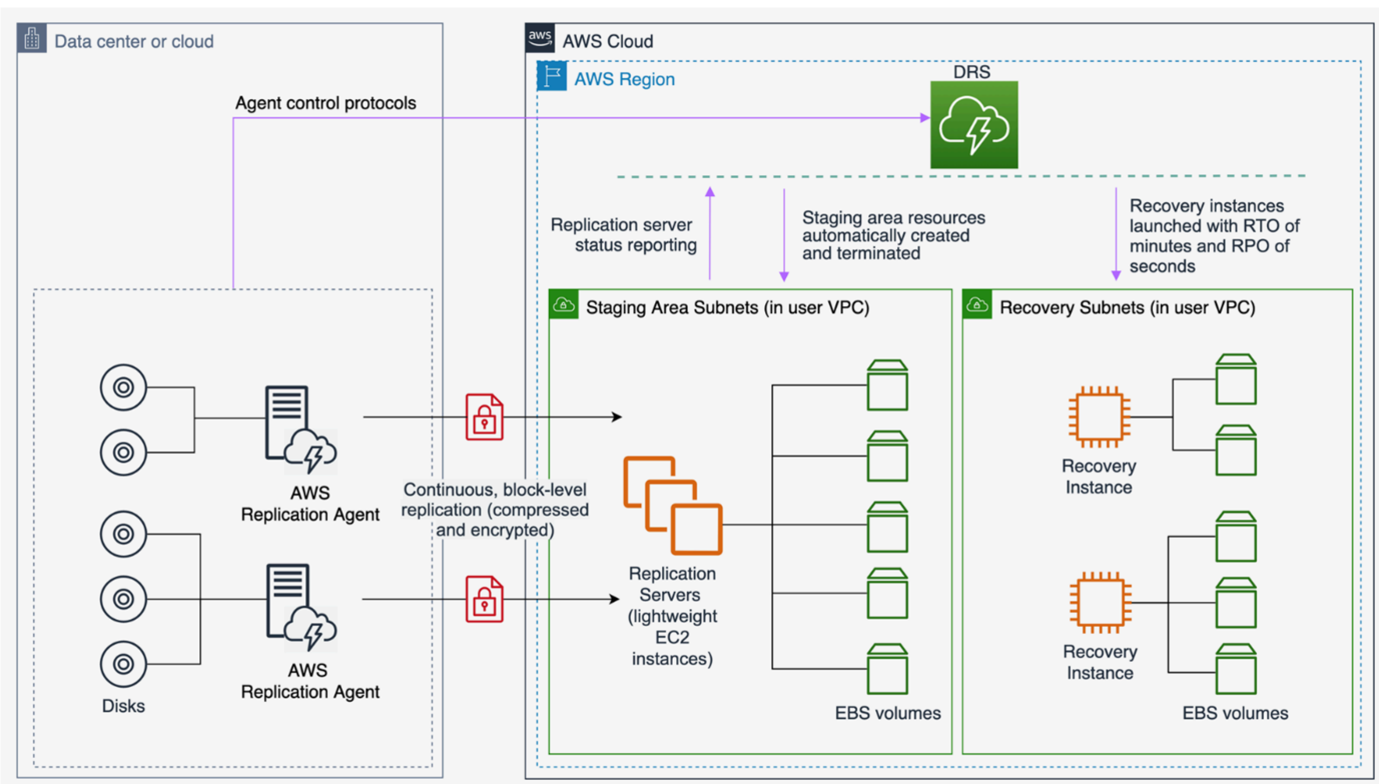

Figura 10 - Architettura AWS Elastic Disaster Recovery

Warm standby

L'approccio warm standby implica la verifica della presenza di una copia ridotta verticalmente, ma comunque funzionale, dell'ambiente di produzione in un'altra regione. Questo approccio estende il concetto di Pilot Light e diminuisce il tempo di ripristino, poiché il carico di lavoro è sempre attivo in un'altra regione. Questo approccio consente inoltre di eseguire più facilmente i test o di implementare test continui per aumentare la fiducia nella capacità di ripristino in caso di emergenza.

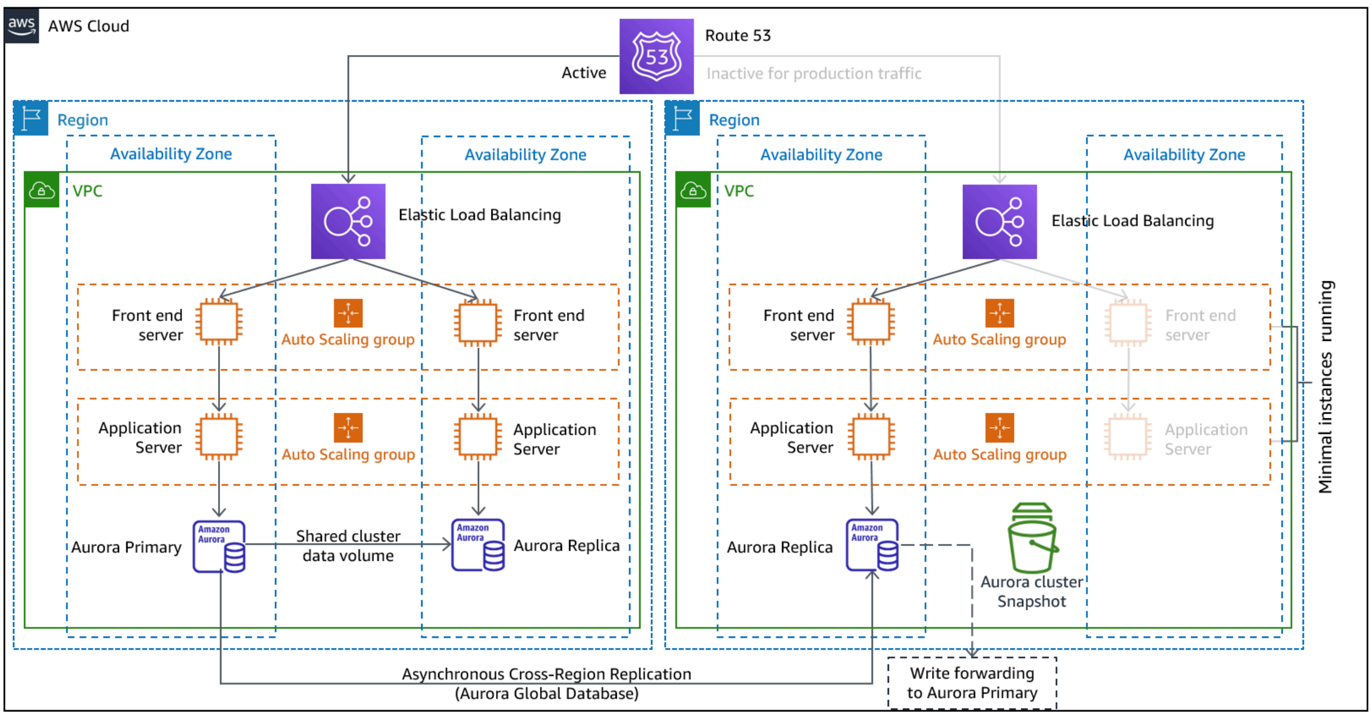

Figura 11 - Architettura Warm Standby

Nota: la differenza tra [luce pilota] e [standby caldo] a volte può essere difficile da capire. Entrambi includono un ambiente nella regione DR con copie delle risorse principali della regione. La differenza è che Pilot Light non può elaborare le richieste senza prima intraprendere un'azione aggiuntiva, mentre lo standby caldo può gestire immediatamente il traffico (a livelli di capacità ridotti). L'approccio pilot light richiede di «accendere» i server, possibilmente implementare un'infrastruttura aggiuntiva (non core) e scalare, mentre lo standby caldo richiede solo la scalabilità (tutto è già installato e funzionante). Utilizzate le vostre esigenze RTO e RPO per aiutarvi a scegliere tra questi approcci.

Servizi AWS

Tutti i servizi AWS coperti da backup e ripristino e pilot light vengono utilizzati anche in modalità warm standby per il backup dei dati, la replica dei dati, il routing active/passive del traffico e l'implementazione dell'infrastruttura, comprese le istanze. EC2

Amazon EC2 Auto Scaling viene utilizzato per scalare risorse tra cui EC2 istanze Amazon, attività Amazon ECS, throughput di Amazon DynamoDB e repliche Amazon Aurora all'interno di una regione AWS. Amazon EC2 Auto Scaling ridimensiona la distribuzione dell' EC2 istanza tra zone di disponibilità all'interno di una regione AWS, fornendo resilienza all'interno di tale regione. Utilizzate Auto Scaling per estendere la vostra regione DR alla piena capacità di produzione, come parte di strategie pilota di standby a luce o a caldo. Ad esempio, per EC2, aumentare l'impostazione di capacità desiderata nel gruppo Auto Scaling. Puoi modificare questa impostazione manualmente tramite Console di gestione AWS, automaticamente tramite l'SDK AWS o ridistribuendo il AWS CloudFormation modello utilizzando il nuovo valore di capacità desiderato. Puoi utilizzare AWS CloudFormation i parametri per semplificare la ridistribuzione del modello. CloudFormation Assicuratevi che le quote di servizio nella vostra regione DR siano sufficientemente alte da non limitare la scalabilità fino alla capacità di produzione.

Poiché l'Auto Scaling è un'attività del piano di controllo, dipendere da essa ridurrà la resilienza della strategia di ripristino complessiva. È un compromesso. È possibile scegliere di fornire una capacità sufficiente in modo che la regione di ripristino sia in grado di gestire l'intero carico di produzione man mano che viene distribuito. Questa configurazione staticamente stabile è denominata hot standby (vedere la sezione successiva). Oppure puoi scegliere di fornire meno risorse, il che ti costerà meno, ma affidandoti all'Auto Scaling. Alcune implementazioni di DR impiegheranno risorse sufficienti per gestire il traffico iniziale, garantendo un RTO basso, e quindi si affideranno all'Auto Scaling per aumentare il traffico successivo.

Attivo/attivo multi-sito

È possibile eseguire il carico di lavoro contemporaneamente in più regioni come parte di una strategia attiva/passiva multisito o attiva/passiva in standby a caldo. Multisito active/active serve il traffico proveniente da tutte le regioni in cui è distribuito, mentre l'hot standby serve il traffico proveniente solo da una singola regione e le altre regioni vengono utilizzate solo per il disaster recovery. Con un active/active approccio multisito, gli utenti sono in grado di accedere al carico di lavoro in qualsiasi regione in cui è distribuito. Questo approccio è l'approccio più complesso e costoso al disaster recovery, ma può ridurre i tempi di ripristino quasi a zero nella maggior parte dei casi di emergenza con le scelte tecnologiche e l'implementazione corrette (tuttavia, il danneggiamento dei dati potrebbe dover fare affidamento sui backup, che di solito portano a un punto di ripristino diverso da zero). L'hot standby utilizza una active/passive configurazione in cui gli utenti vengono indirizzati solo verso una singola regione e le aree DR non assorbono traffico. La maggior parte dei clienti ritiene che, per gestire un ambiente completo nella seconda regione, abbia senso utilizzarlo attivo/attivo. In alternativa, se non si desidera utilizzare entrambe le regioni per gestire il traffico degli utenti, Warm Standby offre un approccio più economico e meno complesso dal punto di vista operativo.

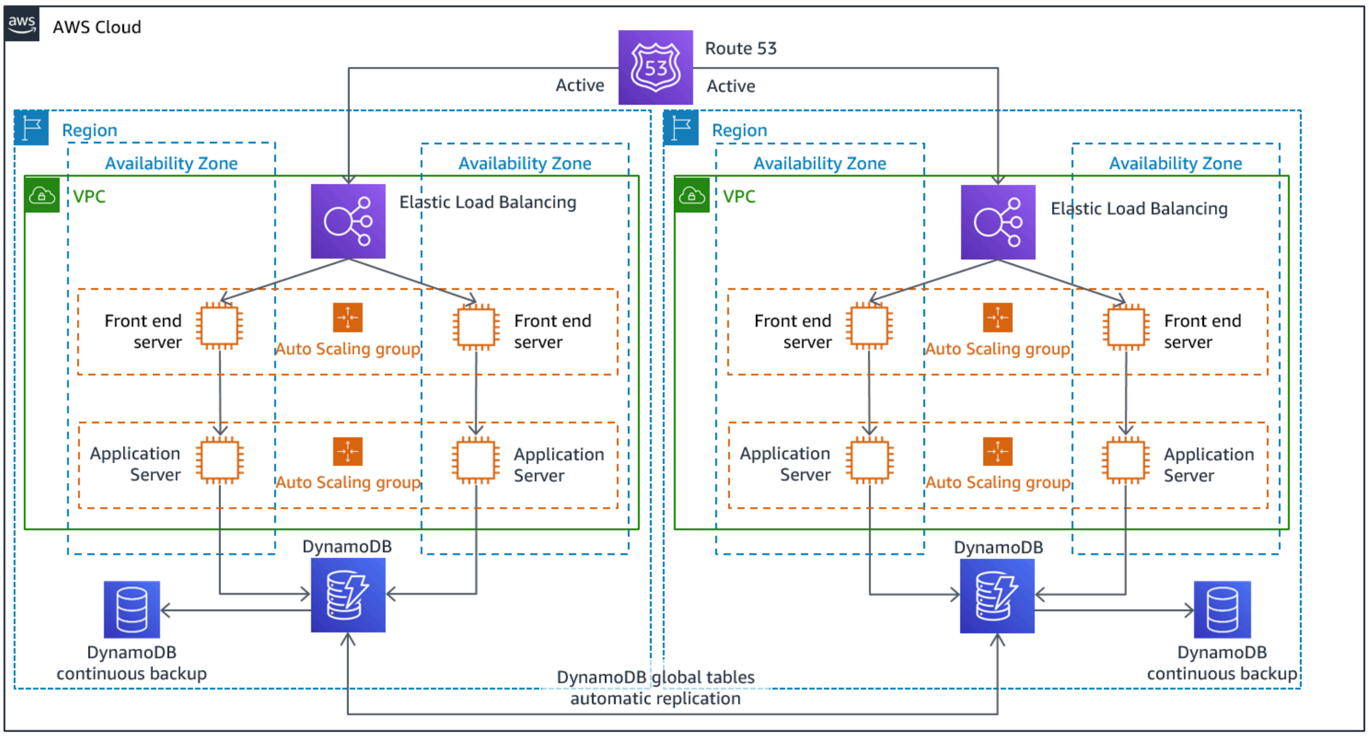

Figura 12 - active/active Architettura multisito (modifica di un percorso attivo in Inattivo per l'hot standby)

Con un approccio multisito active/active, because the workload is running in more than one Region, there is no such thing as failover in this scenario. Disaster recovery testing in this case would focus on how the workload reacts to loss of a Region: Is traffic routed away from the failed Region? Can the other Region(s) handle all the traffic? Testing for a data disaster is also required. Backup and recovery are still required and should be tested regularly. It should also be noted that recovery times for a data disaster involving data corruption, deletion, or obfuscation will always be greater than zero and the recovery point will always be at some point before the disaster was discovered. If the additional complexity and cost of a multi-site active/active (o hot standby) è necessario mantenere tempi di ripristino vicini allo zero, quindi è necessario compiere ulteriori sforzi per mantenere la sicurezza e prevenire gli errori umani per mitigare i disastri umani.

Servizi AWS

Tutti i servizi AWS coperti da backup e ripristino, pilot light e warm standby vengono utilizzati anche qui per il backup point-in-time dei dati, la replica dei dati, il routing active/active del traffico e l'implementazione e la scalabilità dell'infrastruttura, comprese le istanze. EC2

Per gli active/passive scenari descritti in precedenza (Pilot Light e Warm Standby), sia Amazon Route 53 che Amazon AWS Global Accelerator possono essere utilizzati per instradare il traffico di rete verso la regione attiva. Per quanto riguarda la active/active strategia in questo caso, entrambi questi servizi consentono anche la definizione di politiche che determinano quali utenti accedono a quale endpoint regionale attivo. Con l' AWS Global Accelerator utente è possibile impostare una ghiera di controllo del traffico per controllare la percentuale di traffico indirizzata a ciascun endpoint dell'applicazione. Amazon Route 53 supporta questo approccio percentuale e anche molte altre policy disponibili, tra cui quelle basate sulla geoprossimità e sulla latenza. Global Accelerator sfrutta automaticamente l'ampia rete di server edge AWS per effettuare l'onboard del traffico verso la dorsale della rete AWS il prima possibile, con conseguente riduzione delle latenze di richiesta.

La replica asincrona dei dati con questa strategia consente un RPO vicino allo zero. I servizi AWS come il database globale di Amazon Aurora utilizzano un'infrastruttura dedicata che lascia i database completamente disponibili per servire la tua applicazione e possono essere replicati in un massimo di cinque regioni secondarie con una latenza tipica inferiore a un secondo. Sta progettando active/ passive strategies, writes occur only to the primary Region. The difference with active/active il modo in cui viene gestita la coerenza dei dati con le scritture su ciascuna regione attiva. È comune progettare le letture degli utenti in modo che vengano servite dalla regione a loro più vicina, nota come read local. Con le scritture, hai diverse opzioni:

- Una strategia globale di scrittura indirizza tutte le scritture verso una singola regione. In caso di fallimento di quella regione, un'altra regione verrebbe promossa ad accettare le scritture. Il database globale Aurora è ideale per la scrittura globale, in quanto supporta la sincronizzazione con le repliche di lettura tra le regioni ed è possibile promuovere una delle regioni secondarie affinché si assuma read/write le responsabilità in meno di un minuto. Aurora supporta anche l'inoltro di scrittura, che consente ai cluster secondari di un database globale Aurora di inoltrare istruzioni SQL che eseguono operazioni di scrittura al cluster primario.
- Una strategia locale di scrittura indirizza la scrittura nella regione più vicina (proprio come le letture). Le tabelle globali di Amazon DynamoDB abilitano tale strategia, permettendo la lettura e la scrittura da ogni regione in cui viene distribuita la tabella globale. Le tabelle globali di Amazon DynamoDB utilizzano un last writer per la riconciliazione tra aggiornamenti simultanei.

- Una strategia di scrittura partizionata assegna le scritture a una regione specifica in base a una chiave di partizione (come l'ID utente) per evitare conflitti di scrittura. In questo caso è possibile utilizzare la replica di Amazon S3 configurata in modo bidirezionale e attualmente supporta la replica tra due regioni. Quando implementi questo approccio, assicurati di abilitare la sincronizzazione delle modifiche alla replica su entrambi i bucket A e B per replicare le modifiche ai metadati di replica come le liste di controllo degli accessi agli oggetti (ACLs), i tag degli oggetti o i blocchi degli oggetti sugli oggetti replicati. È inoltre possibile configurare se replicare o meno i marker di eliminazione tra i bucket nelle regioni attive. Oltre alla replica, la strategia deve includere anche i point-in-time backup per la protezione da eventi di danneggiamento o distruzione dei dati.

AWS CloudFormation è uno strumento potente per applicare un'infrastruttura distribuita in modo coerente tra gli account AWS in più regioni AWS. AWS CloudFormation StackSetsestende questa funzionalità consentendoti di creare, aggiornare o eliminare CloudFormation stack su più account e regioni con un'unica operazione. Sebbene AWS CloudFormation utilizzi YAML o JSON per definire l'infrastruttura come codice, AWS Cloud Development Kit (AWS CDK)consente di definire l'infrastruttura come codice utilizzando linguaggi di programmazione familiari. Il codice viene convertito in CloudFormation cui viene poi utilizzato per distribuire risorse in AWS.

Rilevamento

È importante sapere il prima possibile che i carichi di lavoro non forniscono i risultati aziendali che avrebbero dovuto fornire. In questo modo, è possibile dichiarare rapidamente un disastro e riprendersi da un incidente. Per obiettivi di ripristino aggressivi, questo tempo di risposta abbinato a informazioni appropriate è fondamentale per raggiungere gli obiettivi di ripristino. Se l'obiettivo del tempo di ripristino è di un'ora, è necessario rilevare l'incidente, informare il personale appropriato, avviare i processi di escalation, valutare le informazioni (se disponibili) sui tempi previsti per il ripristino (senza eseguire il piano di disaster recovery), dichiarare un problema e ripristinare il sistema entro un'ora.

> **Note**
>
> Se le parti interessate decidono di non ricorrere al DR anche se l'RTO sarebbe a rischio, rivalutate i piani e gli obiettivi del DR. La decisione di non invocare i piani di DR può essere dovuta al fatto che i piani sono inadeguati o alla mancanza di fiducia nell'esecuzione.

Nella pianificazione e negli obiettivi è fondamentale tenere conto del rilevamento, della notifica, dell'escalation, dell'individuazione e della dichiarazione degli incidenti per fornire obiettivi realistici e raggiungibili che offrano valore aziendale.

AWS pubblica la maggior parte delle up-to-the-minute informazioni sulla disponibilità dei servizi nel Service Health Dashboard. Controlla in qualsiasi momento per ottenere informazioni aggiornate sullo stato o iscriviti a un feed RSS per ricevere notifiche sulle interruzioni di ogni singolo servizio. Se riscontri un problema operativo in tempo reale con uno dei nostri servizi che non viene visualizzato nella Service Health Dashboard, puoi creare una Support Request.

Dashboard AWS HealthFornisce informazioni sugli AWS Health eventi che possono influire sul tuo account. Le informazioni vengono presentate in due modi: un pannello di controllo che mostra eventi recenti e prossimi organizzati per categoria e un registro completo che mostra tutti gli eventi degli ultimi 90 giorni.

Per soddisfare i requisiti RTO più rigorosi, è possibile implementare il failover automatico basato su controlli di integrità. Progetta controlli di integrità rappresentativi dell'esperienza utente e basati su indicatori chiave di prestazione. I controlli approfonditi dello stato di salute esercitano le funzionalità chiave del carico di lavoro e vanno oltre i controlli superficiali del battito cardiaco. Utilizza controlli sanitari approfonditi basati su più segnali. Utilizzate questo approccio con cautela per non attivare falsi allarmi, poiché il failover quando non è necessario può di per sé comportare rischi di disponibilità.

Test del disaster recovery

Testa l'implementazione del disaster recovery per convalidare l'implementazione e verifica regolarmente il failover nella regione DR del carico di lavoro per assicurarti che RTO e RPO siano soddisfatti.

Un modello da evitare è lo sviluppo di percorsi di ripristino che vengono eseguiti raramente. Ad esempio, è possibile che si disponga di un archivio dati secondario utilizzato per query di sola lettura. Quando scrivi in un archivio dati e quello principale ha un guasto, puoi eseguire il failover verso l'archivio dati secondario. Se non testi frequentemente questo failover, è possibile che i presupposti relativi alle funzionalità dell'archivio dati secondario non siano corretti. La capacità del sistema secondario, che potrebbe essere stata sufficiente durante l'ultimo test, potrebbe non essere più in grado di tollerare il carico in questo scenario oppure le quote di servizio nella regione secondaria potrebbero non essere sufficienti.

La nostra esperienza ha dimostrato che l'unico ripristino da errore che funziona è il percorso sottoposto a frequenti test. Questo è il motivo per cui è preferibile disporre di un numero limitato di percorsi di ripristino.

Puoi stabilire dei modelli di ripristino e testarli regolarmente. Se si dispone di un percorso di ripristino complesso o critico, è comunque necessario eseguire regolarmente tale errore in produzione per verificare che il percorso di ripristino funzioni.

Gestisci le variazioni di configurazione nella regione DR. Assicurati che l'infrastruttura, i dati e la configurazione siano quelli necessari nella regione DR. Ad esempio, verifica che le quote AMIs di servizio siano up-to-date valide.

Puoi utilizzarlo per monitorare e AWS Configregistrare continuamente le configurazioni delle risorse AWS. AWS Config è in grado di rilevare la deriva e attivare AWS Systems Manager Automation per correggere la deriva e generare allarmi. AWS CloudFormationpuò inoltre rilevare la deriva negli stack che hai installato.

Conclusioni

I clienti sono responsabili della disponibilità delle loro applicazioni nel cloud. È importante definire cos'è un disastro e disporre di un piano di disaster recovery che rifletta questa definizione e l'impatto che può avere sui risultati aziendali. Crea Recovery Time Objective (RTO) e Recovery Point Objective (RPO) sulla base dell'analisi dell'impatto e delle valutazioni del rischio, quindi scegli l'architettura appropriata per mitigare i disastri. Garantisci che il rilevamento dei disastri sia possibile e tempestivo: è fondamentale sapere quando gli obiettivi sono a rischio. Assicurati di avere un piano e convalidalo con dei test. I piani di disaster recovery che non sono stati convalidati rischiano di non essere implementati a causa della mancanza di fiducia o del mancato raggiungimento degli obiettivi di disaster recovery.

Collaboratori

Hanno collaborato alla stesura del presente documento:

- Alex Livingstone, responsabile pratico delle operazioni cloud, AWS Enterprise Support
- Seth Eliot, principale architetto di soluzioni di affidabilità, Amazon Web Services

Approfondimenti

Per ulteriori informazioni, consulta:

- AWS Centro di architettura
- Pilastro dell'affidabilità, AWS Well-Architected Framework
- Lista di controllo del piano di disaster recovery
- Implementazione dei controlli sanitari
- Architettura di disaster recovery (DR) su AWS, parte I: Strategie per il ripristino nel cloud
- Architettura di disaster recovery (DR) su AWS, parte II: Backup e ripristino con Rapid Recovery
- Architettura di disaster recovery (DR) su AWS, parte III: Pilot Light and Warm Standby
- Architettura di disaster recovery (DR) su AWS, parte IV: attiva/attiva su più siti
- Creating Disaster Recovery Mechanisms Using Amazon Route 53
- Minimizing Dependencies in a Disaster Recovery Plan
- Laboratori pratici di disaster recovery AWS Well-Architected
- AWS Implementazioni di soluzioni: Multi-Region Application Architecture
- AWS re:Invent 2018: modelli di architettura per applicazioni Active-Active in più regioni (09-R2) ARC2

Cronologia dei documenti

Per ricevere una notifica sugli aggiornamenti del presente whitepaper, iscriviti al feed RSS.

Modifica	Descrizione	Data
Aggiornamenti minori	Correzioni di bug e numerose modifiche minori.	1 aprile 2022
Aggiornamento del whitepaper	Aggiornamenti editoriali minori.	21 marzo 2022
Aggiornamento del whitepaper	Sono state aggiunte informazi oni sul piano dati e sul piano di controllo. Sono stati aggiunti ulteriori dettagli su come implementare il active/passive failover. CloudEndure Disaster Recovery sostituito con AWS Elastic Disaster Recovery.	17 febbraio 2022
Aggiornamento secondario	AWS Well-Architected Tool informazioni aggiunte.	11 febbraio 2022
Pubblicazione iniziale	Whitepaper pubblicato per la prima volta.	12 febbraio 2021

Note

AWS Glossario

Per la AWS terminologia più recente, consultate il AWS glossario nella sezione Reference.Glossario AWS