AWS Whitepaper

AWS Outposts Considerazioni sulla progettazione e sull'architettura ad alta disponibilità

AWS Outposts Considerazioni sulla progettazione e sull'architettura ad alta disponibilità: AWS Whitepaper

Table of Contents

AWS Outposts Considerazioni sulla progettazione e sull'architettura ad alta disponibilità

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

Questo white paper illustra le considerazioni sull'architettura e le pratiche consigliate che i responsabili IT e gli architetti di sistema possono applicare per creare ambienti applicativi locali ad alta disponibilità. AWS Outposts

Sei Well-Architected?

Il 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 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 in AWS Management Console, puoi esaminare i tuoi carichi di lavoro rispetto a queste best practice rispondendo a una serie di domande per ogni pilastro.

Per ulteriori indicazioni e best practice da parte degli esperti per la tua architettura cloud (implementazioni dell'architettura di riferimento, diagrammi e white paper), consulta l'Architecture Center.AWS

Introduzione

Questo paper è destinato ai responsabili IT e agli architetti di sistema che desiderano implementare, migrare e gestire applicazioni utilizzando la piattaforma AWS cloud ed eseguire tali applicazioni on-premise con AWS Outposts rack, il fattore di forma rack 42U di. AWS Outposts

Presenta i modelli di architettura, gli antipattern e le pratiche consigliate per la creazione di sistemi ad alta disponibilità che includono il rack. AWS Outposts Imparerai a gestire la capacità dei AWS Outposts rack e a utilizzare i servizi di rete e delle strutture dei data center per configurare soluzioni di infrastruttura AWS Outposts rack ad alta disponibilità.

AWS Outposts rack è un servizio completamente gestito che fornisce un pool logico di funzionalità di cloud computing, storage e networking. Con i rack Outposts, i clienti possono utilizzare i servizi AWS

gestiti supportati nei loro ambienti locali, tra cui: Amazon Elastic Compute Cloud (Amazon EC2), Amazon Elastic Block Store (AmazonEBS), Amazon S3 su Outposts, Amazon Elastic Kubernetes Service (Amazon EKS), Amazon Elastic Container Service (AmazonECS), Amazon ElasticContainer Service (Amazon ECS), Amazon Relational Amazon Relational Database Service (Amazon RDS) e altri servizi su Outposts.AWS I servizi su Outposts vengono forniti sullo stesso sistema AWS Nitro utilizzato in. Regioni AWS

Sfruttando AWS Outposts rack, puoi creare, gestire e scalare applicazioni locali ad alta disponibilità utilizzando servizi e strumenti AWS cloud familiari. AWS Outposts rack è ideale per carichi di lavoro che richiedono accesso a bassa latenza ai sistemi locali, elaborazione locale dei dati, residenza dei dati e migrazione di applicazioni con interdipendenze di sistema locali.

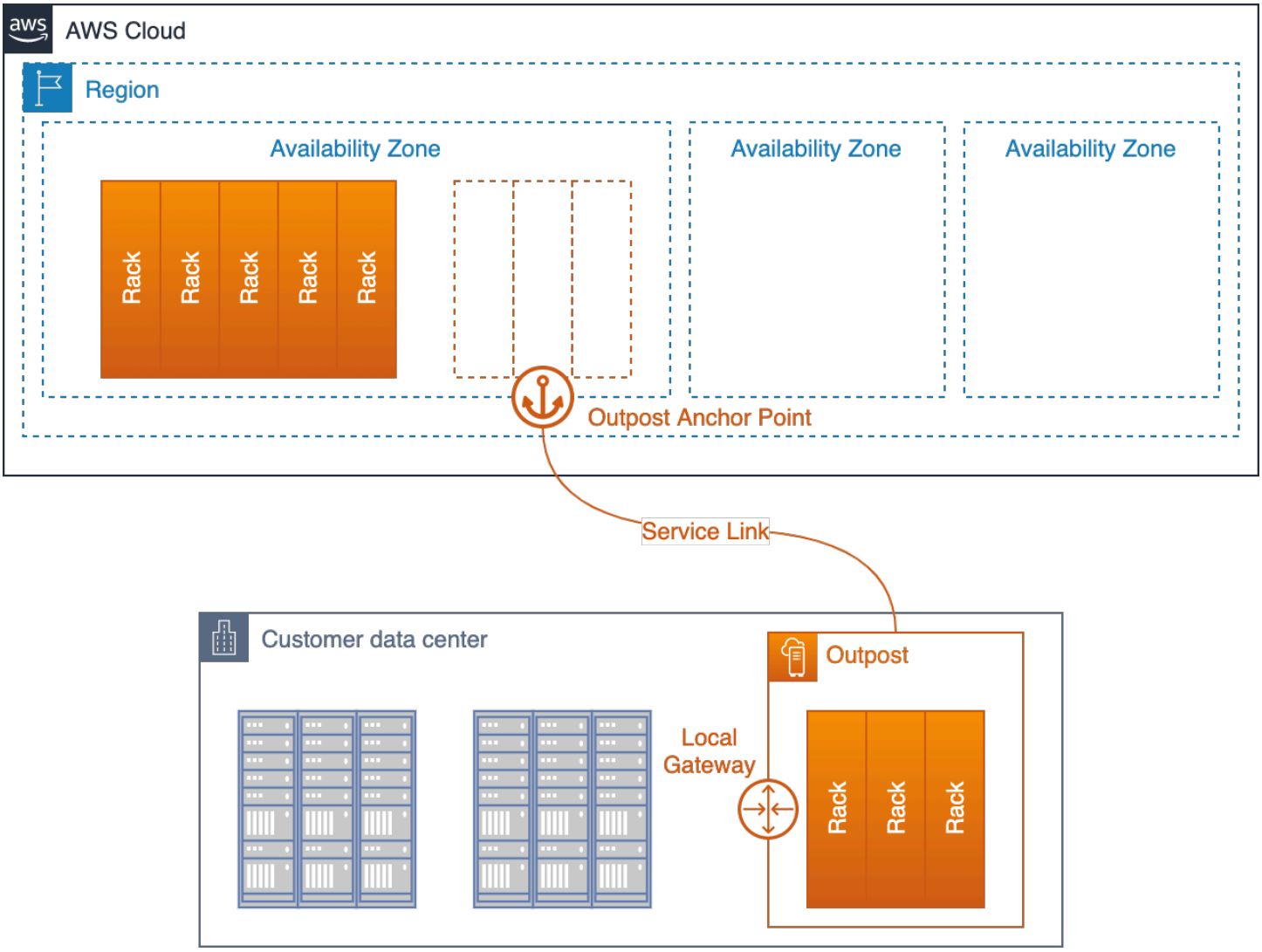
Estensione dell'infrastruttura AWS e dei servizi alle sedi locali

Il AWS Outposts servizio fornisce AWS infrastrutture e servizi a sedi locali in più di 50 paesi e territori, offrendo ai clienti la possibilità di implementare la stessa AWS infrastruttura, gli stessi AWS servizi, API e strumenti praticamente in qualsiasi data center, spazio di co-location o struttura locale per un'esperienza ibrida davvero coerente. Per capire come progettare con Outposts, è necessario comprendere i diversi livelli che compongono il cloud. AWS

An Regione AWSè un'area geografica del mondo. Ciascuno Regione AWS è un insieme di data center raggruppati logicamente in zone di disponibilità (AZ). Regioni AWS forniscono più (almeno due) zone di disponibilità fisicamente separate e isolate, collegate con bassa latenza, throughput elevato e connettività di rete ridondante. Ogni AZ è composta da uno o più data center fisici.

Un Outpost logico (di seguito denominato Outpost) è una distribuzione di uno o più AWS Outposts rack fisicamente collegati e gestiti come un'unica entità. Un Outpost fornisce un pool di capacità di AWS elaborazione e archiviazione in uno dei tuoi siti come estensione privata di una zona di emergenza in un. Regione AWS

Forse il modello concettuale migliore AWS Outposts è quello di pensare di scollegare uno o più rack da un data center in una zona del mondo. Regione AWS Spostate i rack dal data center AZ al vostro data center. Quindi si collegano i rack ai punti di ancoraggio del data center AZ con un cavo (molto) lungo in modo che i rack continuino a funzionare come parte del Regione AWS. Inoltre, li colleghi alla rete locale per fornire connettività a bassa latenza tra le reti locali e i carichi di lavoro in esecuzione su tali rack.



Un Outpost installato in un data center di un cliente e ricollegato alla zona di riferimento AZ e alla regione madre

L'Outpost funge da estensione dell'AZ dove è ancorato. AWS gestisce, monitora e gestisce l' AWS Outposts infrastruttura come parte di. Regione AWS Invece di un cavo fisico molto lungo, un Outpost si ricollega alla regione madre tramite una serie di tunnel VPN crittografati chiamati Service Link.

Il Service Link termina su una serie di punti di ancoraggio in una zona di disponibilità (AZ) nella regione madre dell'Outpost.

Sei tu a scegliere dove archiviare i tuoi contenuti. Puoi replicare ed eseguire il backup dei contenuti nella Regione AWS o in altre posizioni. I tuoi contenuti non verranno spostati o copiati al di fuori delle località prescelte senza il tuo consenso, a meno che non sia necessario per rispettare la legge o un

ordine vincolante di un ente governativo. Per ulteriori informazioni, consulta le Domande frequenti sulla privacy dei dati in AWS.

I carichi di lavoro distribuiti su questi rack vengono eseguiti localmente. Inoltre, sebbene la capacità di elaborazione e archiviazione disponibile in questi rack sia limitata e non sia in grado di supportare l'esecuzione dei servizi su scala cloud di un utente Regione AWS, le risorse distribuite sul rack (le istanze e lo storage locale) ottengono i vantaggi dell'esecuzione locale mentre il piano di gestione continua a funzionare in. Regione AWS

Per distribuire carichi di lavoro su un Outpost, aggiungi sottoreti ai tuoi ambienti Virtual Private Cloud (VPC) e specifica un Outpost come posizione per le sottoreti. Quindi, si seleziona la sottorete desiderata quando si distribuiscono AWS le risorse supportate tramite gli strumenti AWS Management Console CLI, API, CDK o Infrastructure as Code (IaC). Le istanze nelle sottoreti Outpost comunicano con altre istanze sull'Outpost o nella regione tramite reti VPC.

Outpost Service Link trasporta sia il traffico di gestione di Outpost che il traffico VPC dei clienti (traffico VPC tra le sottoreti di Outpost e le sottoreti nella regione).

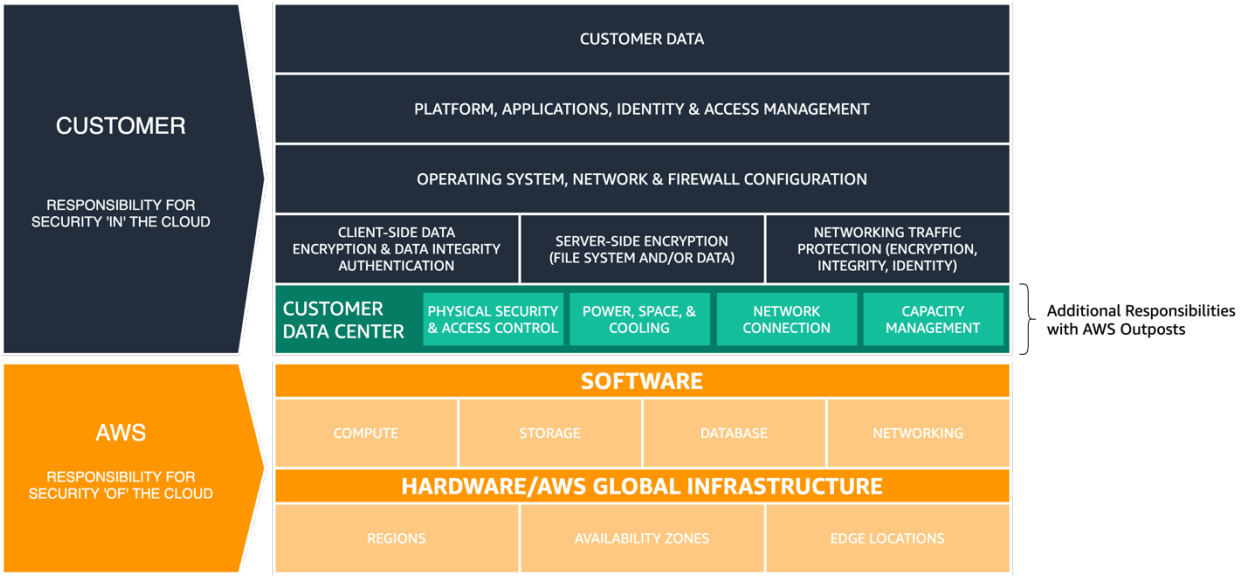
Termini importanti:

- AWS Outposts— è un servizio completamente gestito che offre la stessa AWS infrastruttura, gli stessi AWS servizi, API e strumenti praticamente a qualsiasi data center, spazio di co-location o struttura locale per un'esperienza ibrida davvero coerente.
- Outpost: è un'implementazione di uno o più AWS Outposts rack fisicamente connessi che viene gestita come un'unica entità logica e un pool di AWS elaborazione, storage e rete distribuiti presso la sede del cliente.
- Regione principale: la regione Regione AWS che fornisce i servizi di gestione, piano di controllo e i AWS servizi regionali per un'implementazione Outpost.
- Zona di disponibilità di ancoraggio (anchor AZ): la zona di disponibilità nella regione madre che ospita i punti di ancoraggio di un avamposto. Un avamposto funge da estensione della sua zona di disponibilità di ancoraggio.
- Punti di ancoraggio: endpoint nell'anchor AZ che ricevono le connessioni dagli Outposts distribuiti in remoto.
- Service Link: un insieme di tunnel VPN crittografati che collegano un Outpost alla sua zona di disponibilità di ancoraggio nella regione madre.
- Local Gateway (LGW): un router virtuale di interconnessione logica che consente la comunicazione tra Outpost e la rete locale.

Comprensione del modello di responsabilità condivisa aggiornato

Quando si implementa l' AWS Outposts infrastruttura nei data center o nelle strutture di co-ubicazione, si assumono responsabilità aggiuntive nel modello di responsabilitàAWS condivisa. Ad esempio, nella regione, AWS fornisce diverse fonti di alimentazione, reti di base ridondanti e connettività WAN (Wide Area Network) resiliente per garantire la disponibilità dei servizi in caso di guasto di uno o più componenti.

Con Outposts, sei responsabile di fornire alimentazione e connettività di rete resilienti ai rack Outpost per soddisfare i requisiti di disponibilità per i carichi di lavoro in esecuzione su Outposts.



AWS Modello di responsabilità condivisa aggiornato per AWS Outposts

Con AWS Outposts, sei responsabile della sicurezza fisica e del controllo degli accessi dell'ambiente del data center. È necessario fornire alimentazione, spazio e raffreddamento sufficienti per mantenere operativo l'Outpost e le connessioni di rete per ricollegare l'Outpost alla Regione.

Poiché la capacità di Outpost è limitata e determinata dalle dimensioni e dal numero di rack AWS installati sul tuo sito, devi decidere quanta capacità di EC2, EBS e S3 on Outposts ti serve per eseguire i carichi di lavoro iniziali, far fronte alle crescite future e fornire capacità aggiuntiva per mitigare i guasti dei server e gli eventi di manutenzione.

AWS è responsabile della disponibilità dell'infrastruttura Outposts, compresi gli alimentatori, i server e le apparecchiature di rete all'interno dei AWS Outposts rack. AWS gestisce anche l'hypervisor di virtualizzazione, i sistemi di storage e i AWS servizi eseguiti su Outposts.

Un power shelf centrale in ogni rack Outposts converte l'alimentazione da AC a DC e fornisce alimentazione ai server nel rack tramite un'architettura bus bar. Con l'architettura bus bar, metà degli alimentatori del rack possono guastarsi e tutti i server continueranno a funzionare senza interruzioni.

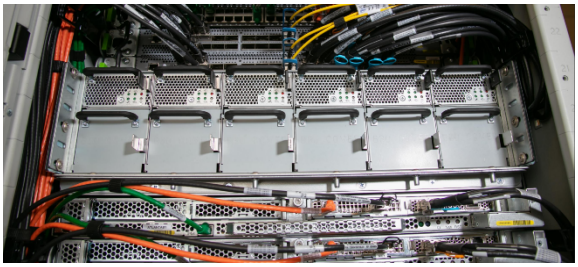


Figura 3 - Alimentatori AWS Outposts AC-DC e distribuzione dell'alimentazione tramite bus bar

Anche gli switch di rete e i cavi all'interno e tra i rack Outposts sono completamente ridondanti. Un patch panel in fibra fornisce la connettività tra un rack Outpost e la rete locale e funge da punto di demarcazione tra l'ambiente del data center gestito dal cliente e l'ambiente gestito. AWS Outposts

Proprio come nella regione, AWS è responsabile dei servizi cloud offerti su Outposts e si assume responsabilità aggiuntive man mano che selezioni e distribuisci servizi gestiti di livello superiore come Amazon RDS on Outposts. È necessario consultare il Modello di responsabilitàAWS condivisa e le pagine delle domande frequenti (FAQ) per i singoli servizi mentre consideri e selezioni i servizi da distribuire su Outposts. Queste risorse forniscono dettagli aggiuntivi sulla divisione delle responsabilità tra l'utente e. AWS

Pensare in termini di modalità di fallimento

Quando si progetta un'applicazione o un sistema ad alta disponibilità, è necessario considerare quali componenti potrebbero guastarsi, quale impatto avranno i guasti dei componenti sul sistema e quali meccanismi è possibile implementare per mitigare o eliminare l'impatto dei guasti dei componenti. L'applicazione viene eseguita su un singolo server, in un singolo rack o in un singolo data center? Cosa succede quando un server, un rack o un data center subisce un guasto temporaneo o permanente? Cosa succede in caso di guasto in un sottosistema critico come la rete o all'interno dell'applicazione stessa? Si tratta di modalità di errore.

È necessario considerare le modalità di errore in questa sezione quando si pianificano gli Outposts e le distribuzioni delle applicazioni. Le sezioni che seguono esamineranno come mitigare queste modalità di errore per fornire un maggiore livello di alta disponibilità per l'ambiente applicativo.

Modalità di errore 1: rete

Una distribuzione di Outpost dipende da una connessione resiliente alla regione madre per la gestione e il monitoraggio. Le interruzioni della rete possono essere causate da una serie di guasti, come errori dell'operatore, guasti delle apparecchiature e interruzioni dei fornitori di servizi. Un avamposto, che può essere composto da uno o più rack collegati tra loro nel sito, è considerato disconnesso quando non è in grado di comunicare con la regione tramite il Service Link.

I percorsi di rete ridondanti possono contribuire a ridurre il rischio di eventi di disconnessione. È necessario mappare le dipendenze delle applicazioni e il traffico di rete per comprendere l'impatto degli eventi di disconnessione sulle operazioni del carico di lavoro. Pianifica una ridondanza di rete sufficiente a soddisfare i requisiti di disponibilità delle applicazioni.

Durante un evento di disconnessione, le istanze in esecuzione su Outpost continuano a funzionare e sono accessibili dalle reti locali tramite Outpost Local Gateway (LGW). I carichi di lavoro e i servizi locali potrebbero essere compromessi o fallire se si affidano ai servizi della regione. Le richieste mutevoli (ad esempio l'avvio o l'arresto di istanze su Outpost), le operazioni del piano di controllo e la telemetria del servizio (ad esempio, le CloudWatch metriche) avranno esito negativo quando l'Outpost è disconnesso dalla Regione.

Modalità di errore 2: Istanze

Le istanze EC2 possono danneggiarsi o fallire se il server su cui sono in esecuzione presenta un problema o se l'istanza presenta un errore del sistema operativo o dell'applicazione. Il modo in cui le

applicazioni gestiscono questi tipi di errori dipende dall'architettura dell'applicazione. Le applicazioni monolitiche utilizzano in genere le funzionalità dell'applicazione o del sistema per il ripristino, mentre le architetture modulari orientate ai servizi o ai microservizi in genere sostituiscono i componenti guasti per mantenere la disponibilità del servizio.

Puoi sostituire le istanze fallite con nuove istanze utilizzando meccanismi automatici come i gruppi di Auto Scaling EC2. Il ripristino automatico delle istanze può riavviare le istanze che falliscono a causa di guasti del server, a condizione che sia disponibile una capacità di riserva sufficiente sui server rimanenti.

Modalità di errore 3: calcolo

I server possono guastarsi o danneggiarsi e potrebbe essere necessario metterli fuori servizio (temporaneamente o permanentemente) per una serie di motivi, ad esempio guasti dei componenti e operazioni di manutenzione programmata. Il modo in cui i servizi sul rack Outposts gestiscono i guasti e i problemi del server varia e può dipendere dal modo in cui i clienti configurano le opzioni di alta disponibilità.

È necessario ordinare una capacità di elaborazione sufficiente a supportare un modello di `N+M` disponibilità, indicando qual `N` è la capacità richiesta e quella di riserva fornita per far `M` fronte ai guasti del server.

Le sostituzioni hardware per i server guasti vengono fornite come parte del servizio rack completamente gestito. AWS Outposts AWS monitora attivamente lo stato di tutti i server e i dispositivi di rete in una distribuzione Outpost. Se è necessario eseguire la manutenzione fisica, AWS pianificherà una visita al sito per sostituire i componenti guasti. Il provisioning di capacità inutilizzata consente di mantenere attivi i carichi di lavoro mentre i server guasti vengono messi fuori servizio e sostituiti.

Modalità di errore 4: rack o data center

I guasti ai rack possono verificarsi a causa della perdita totale di alimentazione dei rack o a causa di problemi ambientali come la perdita del raffreddamento o danni fisici al data center a causa di alluvioni o terremoti. Le carenze nelle architetture di distribuzione dell'alimentazione dei centri dati o gli errori durante la manutenzione standard dell'alimentazione dei centri dati possono causare la perdita di alimentazione di uno o più rack o addirittura dell'intero data center.

Questi scenari possono essere mitigati implementando l'infrastruttura su più piani o sedi di data center indipendenti l'una dall'altra all'interno dello stesso campus o area metropolitana.

L'adozione di questo approccio con il AWS Outposts rack richiederà un'attenta considerazione del modo in cui le applicazioni sono architettate e distribuite per funzionare su più Outpost logici separati per mantenere la disponibilità delle applicazioni.

Modalità di errore 5: zona o regione di AWS disponibilità

Ogni avamposto è ancorato a una zona di disponibilità (AZ) specifica all'interno di un. Regione AWS I guasti all'interno della zona di ancoraggio AZ o della regione madre potrebbero causare la perdita della gestione e della mutabilità dell'avamposto e potrebbero interrompere le comunicazioni di rete tra l'avamposto e la regione.

Analogamente ai guasti della rete, i guasti della zona o della regione possono causare la disconnessione dell'Outpost dalla regione. Le istanze in esecuzione su Outpost continuano a funzionare e sono accessibili dalle reti locali tramite Outpost Local Gateway (LGW) e potrebbero essere danneggiate o fallire se si affidano ai servizi della Regione, come descritto in precedenza.

Per mitigare l'impatto dei guasti in AWS AZ e Region, puoi implementare più Outposts, ciascuno ancorato a una zona o regione diversa. È quindi possibile progettare il carico di lavoro in modo che operi in un modello di implementazione distribuito Multi-Outpost utilizzando molti dei meccanismi e dei modelli architettonici simili utilizzati oggi per la progettazione e l'implementazione. AWS

Creazione di applicazioni HA e soluzioni di infrastruttura con AWS Outposts rack

Con AWS Outposts rack, puoi creare, gestire e scalare applicazioni locali ad alta disponibilità utilizzando servizi e strumenti cloud familiari. AWS È importante comprendere che le architetture e gli approcci al cloud HA sono generalmente diversi dalle tradizionali architetture HA locali che potresti utilizzare oggi nel tuo datacenter.

Con le tradizionali implementazioni di applicazioni HA locali, le applicazioni vengono distribuite in macchine virtuali (VM). I sistemi e le infrastrutture IT complessi vengono implementati e mantenuti per mantenere tali macchine virtuali funzionanti e integre. Le macchine virtuali hanno spesso identità specifiche e ciascuna macchina virtuale può svolgere un ruolo fondamentale nell'architettura totale dell'applicazione.

I ruoli architettonici sono strettamente associati alle identità delle VM. Gli architetti di sistema sfruttano le funzionalità dell'infrastruttura IT per fornire ambienti di runtime VM ad alta disponibilità che forniscono a ciascuna macchina virtuale un accesso affidabile alla capacità di elaborazione, ai volumi di storage e ai servizi di rete. In caso di guasto di una macchina virtuale, vengono eseguiti processi di ripristino automatici o manuali per ripristinare lo stato integro della macchina virtuale guasta, spesso su un'altra infrastruttura o interamente in un altro datacenter.

Le architetture Cloud HA adottano un approccio diverso. AWS i servizi cloud forniscono funzionalità di elaborazione, archiviazione e rete affidabili. I componenti dell'applicazione vengono distribuiti su istanze EC2, contenitori, funzioni serverless o altri servizi gestiti.

Un'istanza è un'istanza di un componente dell'applicazione, forse uno dei tanti che svolgono quel ruolo. I componenti dell'applicazione sono strettamente associati tra loro e al ruolo che svolgono nell'architettura totale dell'applicazione. L'identità individuale di un'istanza in genere non è importante. È possibile creare o distruggere istanze aggiuntive per aumentare o ridurre in base alla domanda. Le istanze non riuscite o non integre vengono semplicemente sostituite con nuove istanze integre.

AWS Outposts rack è un servizio completamente gestito che estende AWS elaborazione, archiviazione, rete, database e altri servizi cloud a sedi locali per un'esperienza ibrida davvero coerente. Non dovresti pensare al servizio rack Outposts come a un sostituto immediato dei sistemi di infrastruttura IT con meccanismi HA tradizionali locali. Il tentativo di utilizzare AWS servizi e Outposts per supportare un'architettura HA tradizionale on-premise è un anti-pattern.

I carichi di lavoro in esecuzione su AWS Outposts rack utilizzano meccanismi cloud HA come Amazon EC2 Auto Scaling (per scalare orizzontalmente per soddisfare le richieste del carico di lavoro), i controlli dello stato di EC2 (per rilevare e rimuovere le istanze non integre) e gli Application Load Balancer (per reindirizzare il traffico dei carichi di lavoro in entrata verso istanze scalate o sostituite). Durante la migrazione delle applicazioni sul cloud, sia su un rack che su un Regione AWS AWS Outposts rack, è necessario aggiornare l'architettura dell'applicazione HA per iniziare a sfruttare i servizi cloud gestiti e i meccanismi di disponibilità elevata del cloud.

Le sezioni seguenti introducono modelli di architettura, anti-pattern e pratiche consigliate per l'implementazione di AWS Outposts rack negli ambienti locali per eseguire carichi di lavoro con requisiti di elevata disponibilità. Queste sezioni introducono modelli e pratiche; tuttavia, non forniscono dettagli di configurazione e implementazione. È necessario leggere e acquisire familiarità con le domande frequenti e la Guida per l'utente di AWS Outposts rack e le domande frequenti e la documentazione di servizio per i servizi eseguiti sul rack Outposts mentre prepari l'ambiente per il rack Outposts e le applicazioni per la migrazione ai servizi. AWS

Argomenti

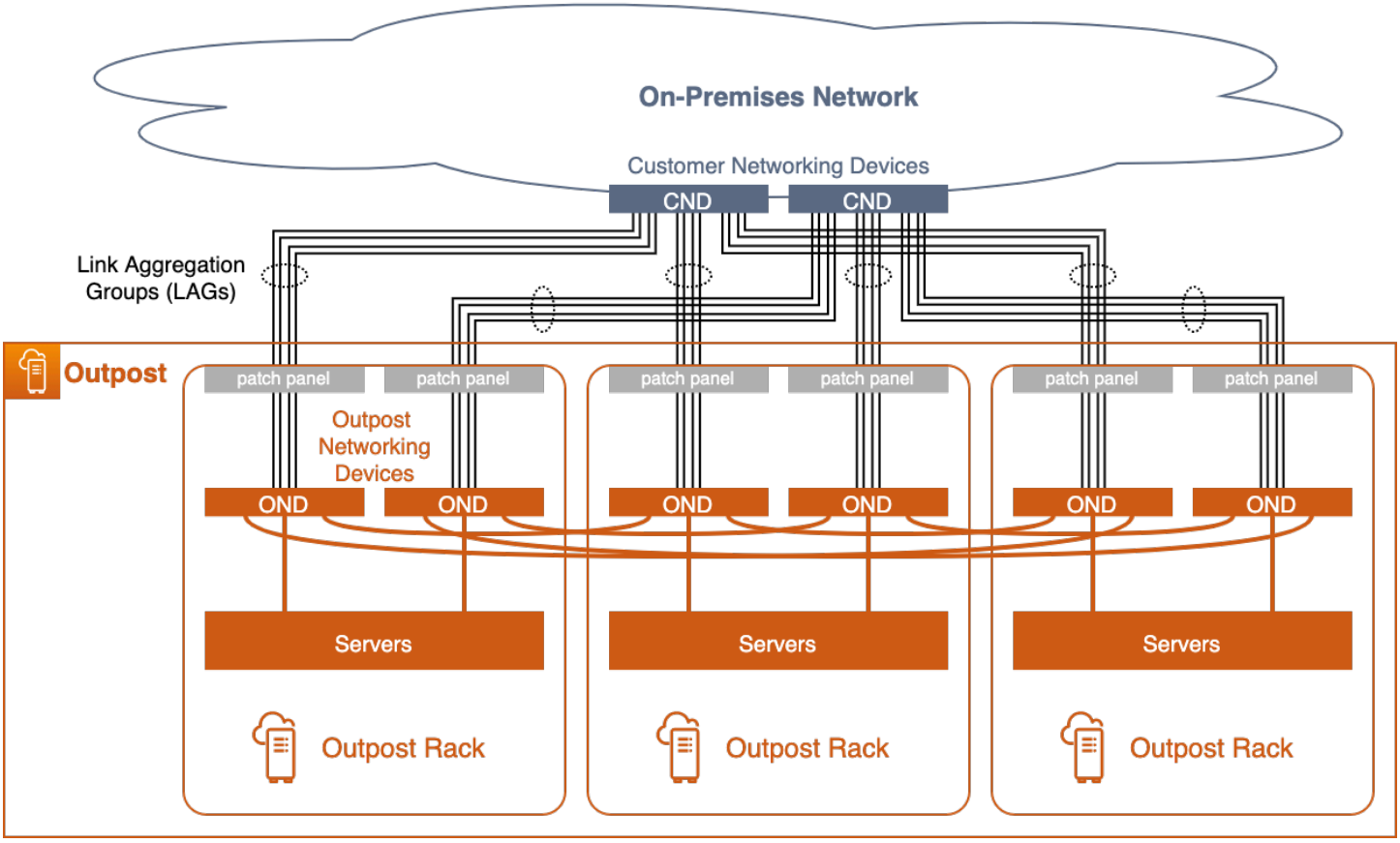
Rete

Per il corretto funzionamento delle operazioni di gestione, monitoraggio e assistenza, l'implementazione di Outpost dipende da una connessione resiliente alla relativa area di riferimento AZ. È necessario effettuare il provisioning della rete locale per fornire connessioni di rete ridondanti per ogni rack Outpost e una connettività affidabile verso i punti di ancoraggio nel cloud. AWS Considera anche i percorsi di rete tra i carichi di lavoro delle applicazioni in esecuzione su Outpost e gli altri sistemi locali e cloud con cui comunicano: come indirizzerai questo traffico nella tua rete?

Argomenti

Allegato di rete

Ogni AWS Outposts rack è configurato con top-of-rack switch ridondanti denominati Outpost Networking Devices (OND). I server di elaborazione e archiviazione in ogni rack si connettono a entrambi gli ODN. È necessario collegare ogni OND a uno switch separato chiamato Customer Networking Device (CND) nel data center per fornire percorsi fisici e logici diversi per ogni rack Outpost. Le OND si collegano ai CND con una o più connessioni fisiche utilizzando cavi in fibra ottica e ricetrasmettitori ottici. Le connessioni fisiche sono configurate in collegamenti LAG (Logical Link Aggregation Group).



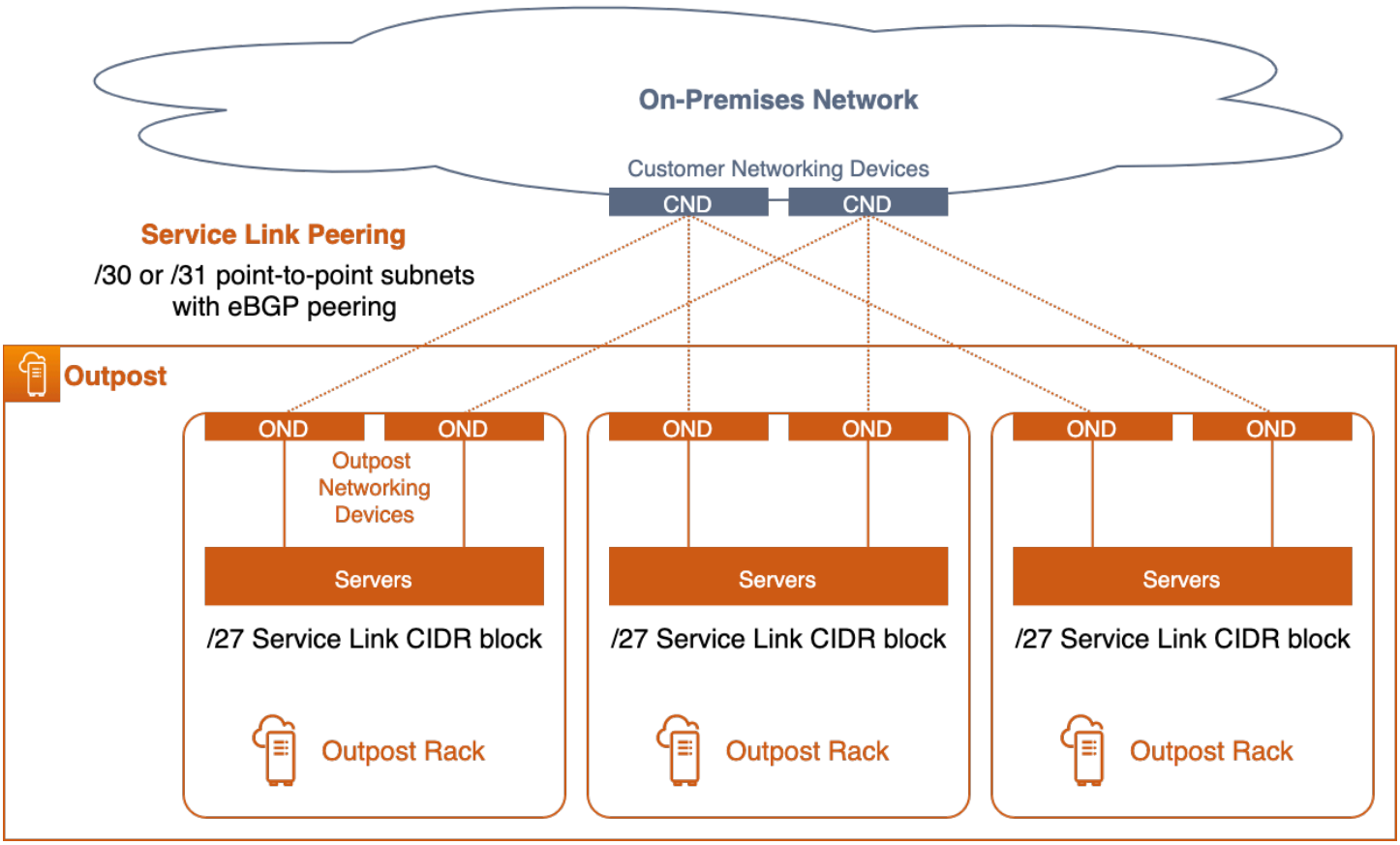
Multi-rack Outpost con allegati di rete ridondanti

I collegamenti da OND a CND sono sempre configurati in un LAG, anche se la connessione fisica è un singolo cavo in fibra ottica. La configurazione dei collegamenti come gruppi LAG consente di aumentare la larghezza di banda del collegamento aggiungendo ulteriori connessioni fisiche al gruppo logico. I link LAG sono configurati come trunk Ethernet IEEE 802.1q per consentire reti separate tra Outpost e la rete locale.

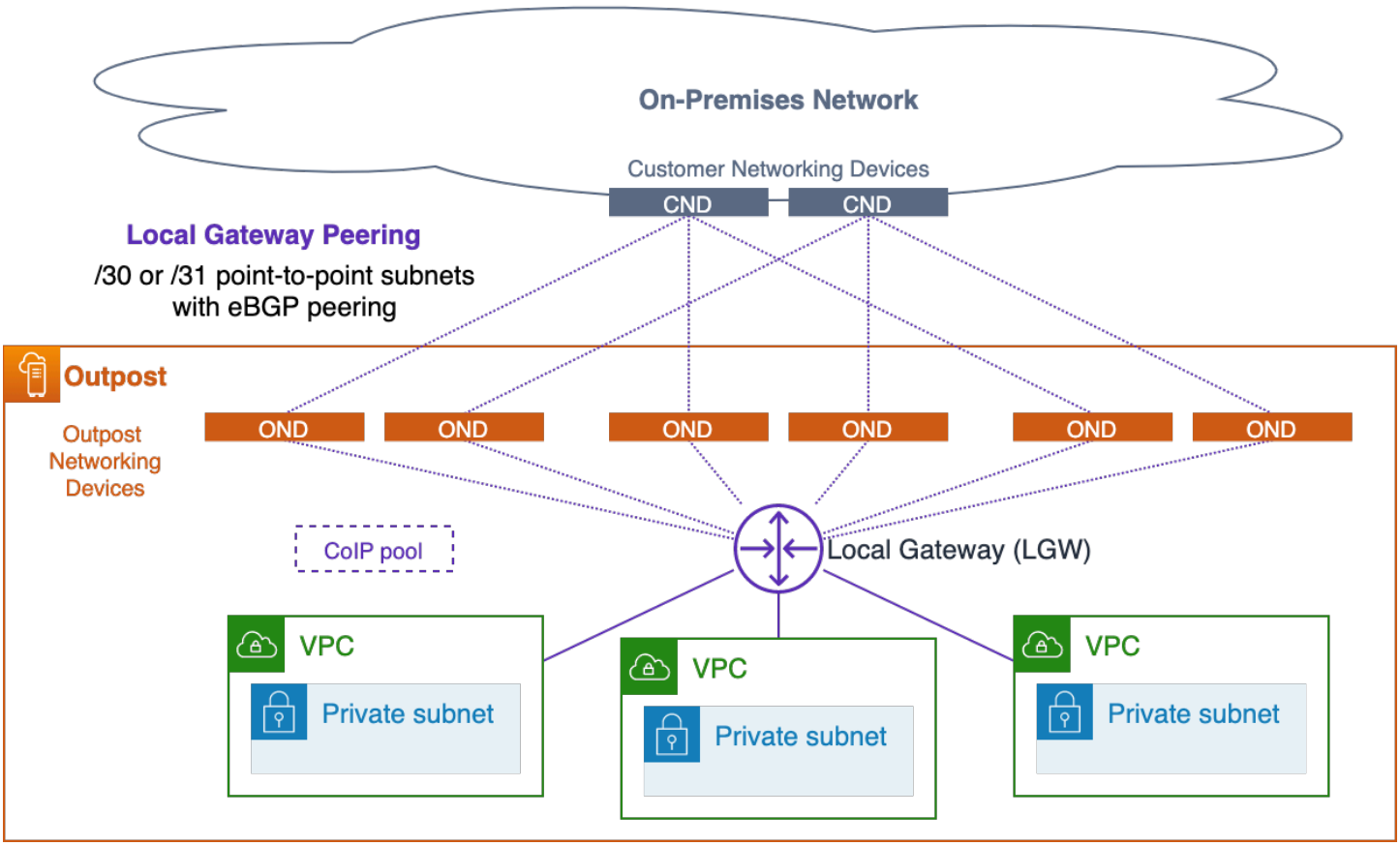
Ogni Outpost dispone di almeno due reti logicamente separate che devono comunicare con o attraverso la rete dei clienti:

- Service link network: alloca gli indirizzi IP del service link ai server Outpost e facilita la comunicazione con la rete locale per consentire ai server di riconnettersi ai punti di ancoraggio Outpost nella regione.
- Rete Gateway locale: consente la comunicazione tra le sottoreti VPC su Outpost e la rete locale tramite Outpost Local Gateway (LGW).

Queste reti separate si collegano alla rete locale tramite una serie di connessioni IP tramite i collegamenti LAG. point-to-point Ogni collegamento LAG da OND a CND è configurato con ID VLAN, sottoreti IP point-to-point (/30 o /31) e peering eBGP per ogni rete separata (service link e LGW). È necessario considerare i collegamenti LAG, con le relative VLAN e sottoreti, come connessioni di livello 2 segmentate e instradate di livello 3. point-to-point Le connessioni IP instradate forniscono percorsi logici ridondanti che facilitano la comunicazione tra le reti separate su Outpost e la rete locale.



Peering dei collegamenti di servizio



Peering del gateway locale

È necessario interrompere i collegamenti LAG di livello 2 (e le relative VLAN) sugli switch CND collegati direttamente e configurare le interfacce IP e il peering BGP sugli switch CND. Non è necessario collegare le VLAN LAG tra gli switch del data center. Per ulteriori informazioni, consulta Connettività a livello di rete nella Guida per l'utente.AWS Outposts

All'interno di un Outpost logico multirack, gli OD sono interconnessi in modo ridondante per fornire connettività di rete ad alta disponibilità tra i rack e i carichi di lavoro in esecuzione sui server. AWS è responsabile della disponibilità della rete all'interno dell'Outpost.

Pratiche consigliate per un collegamento di rete ad alta disponibilità

- Connect ogni Outpost Networking Device (OND) in un rack Outpost a un Customer Networking Device (CND) separato nel data center.
- Interrompi i collegamenti di livello 2, le VLAN, le sottoreti IP di livello 3 e il peering BGP sugli switch CND (Customer Networking Device) collegati direttamente. Non collegate le VLAN OND a CND tra i CND o attraverso la rete locale.

- Aggiungi collegamenti ai Link Aggregation Groups (LAG) per aumentare la larghezza di banda disponibile tra Outpost e il data center. Non fate affidamento sulla larghezza di banda aggregata dei diversi percorsi che attraversano entrambi gli ODN.
- Utilizza i diversi percorsi attraverso le ODN ridondanti per fornire una connettività resiliente tra le reti Outpost e la rete locale.
- Per ottenere una ridondanza ottimale e consentire una manutenzione OND senza interruzioni, consigliamo ai clienti di configurare gli annunci e le politiche BGP come segue:
  - Le apparecchiature di rete del cliente devono ricevere annunci BGP da Outpost senza modificare gli attributi BGP e consentire il multipath/load-balancing BGP per ottenere flussi di traffico in entrata ottimali (dal cliente verso Outpost). La prependenza AS-Path viene utilizzata per i prefissi BGP di Outpost per spostare il traffico da un particolare OND/uplink nel caso in cui sia necessaria una manutenzione. La rete di clienti dovrebbe preferire percorsi provenienti da Outpost con lunghezza AS-Path 1 rispetto a percorsi con AS-Path lunghezza 4, ovvero reagire alla prependenza di AS-Path.
  - La rete di clienti dovrebbe pubblicizzare prefissi BGP uguali con gli stessi attributi per tutti gli ODN di Outpost. Per impostazione predefinita, la rete Outpost bilancia il carico del traffico in uscita (verso il cliente) tra tutti gli uplink. Le politiche di routing vengono utilizzate sul lato Outpost per spostare il traffico lontano da un particolare OND nel caso in cui sia necessaria una manutenzione. Per eseguire questo spostamento del traffico ed eseguire la manutenzione senza interruzioni, sono necessari prefissi BGP uguali da parte del cliente su tutti gli ODN. Quando è necessaria la manutenzione della rete del cliente, consigliamo di utilizzare AS-Path Prepending per allontanare temporaneamente il traffico da un particolare uplink o dispositivo.

Connettività Anchor

Un collegamento al servizio Outpost si collega a punti di ancoraggio pubblici o privati (non entrambi) in una zona di disponibilità (AZ) specifica nella regione madre di Outpost. I server Outpost avviano le connessioni VPN di service link in uscita dai rispettivi indirizzi IP di service link ai punti di ancoraggio nell'anchor AZ. Queste connessioni utilizzano le porte UDP e TCP 443. AWS è responsabile della disponibilità dei punti di ancoraggio nella Regione.

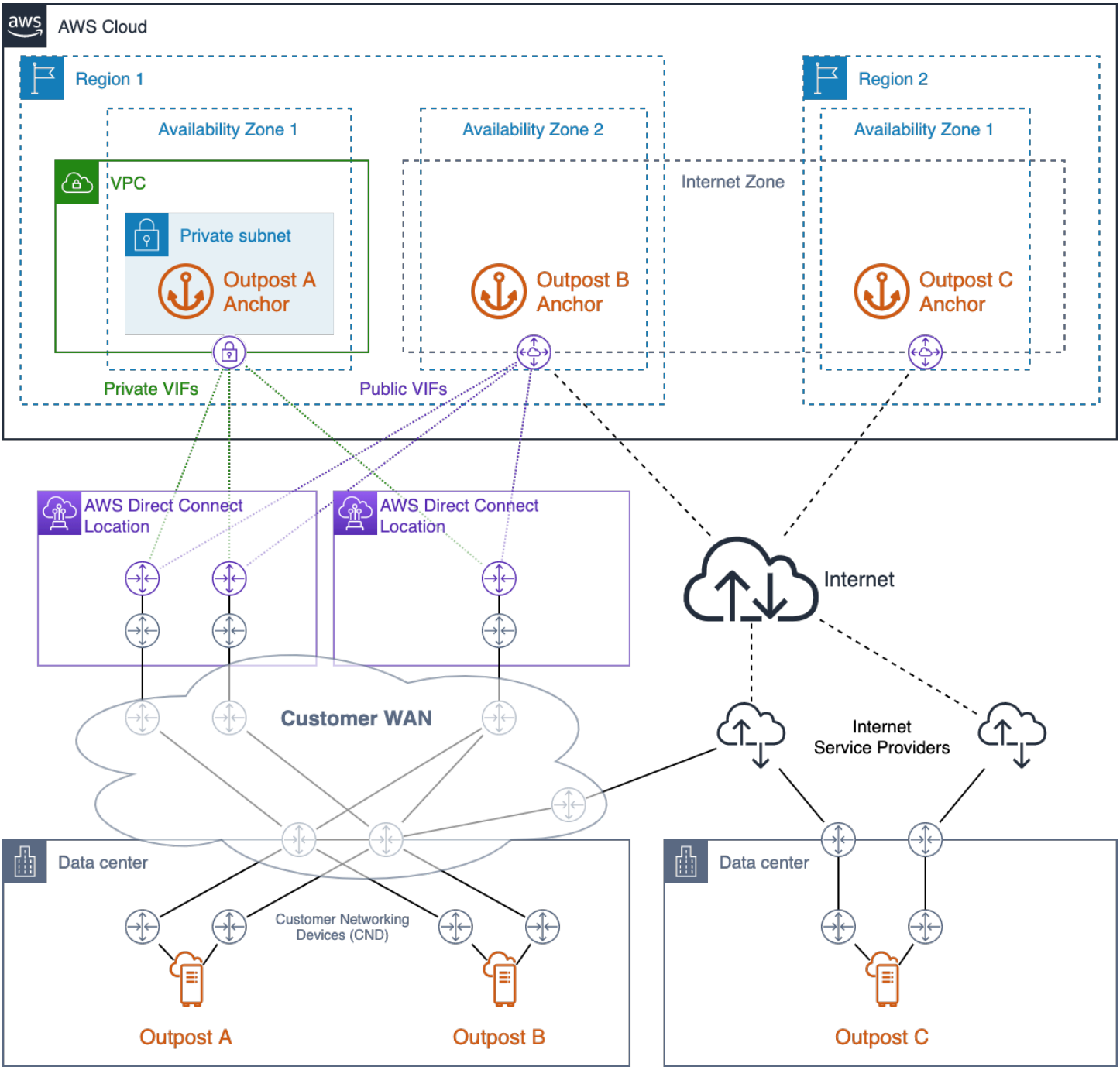
È necessario assicurarsi che gli indirizzi IP del servizio Outpost siano in grado di connettersi attraverso la rete ai punti di ancoraggio dell'Anchor AZ. Gli indirizzi IP del service link non devono comunicare con altri host sulla rete locale.

I punti di ancoraggio pubblici risiedono negli intervalli IP pubblici della regione (nei blocchi CIDR del servizio EC2) e sono accessibili tramite Internet o tramite interfacce virtuali pubbliche (VIF) AWS Direct Connect(DX). L'uso di punti di ancoraggio pubblici consente una selezione più flessibile del percorso in quanto il traffico dei collegamenti di servizio può essere instradato su qualsiasi percorso disponibile in grado di raggiungere con successo i punti di ancoraggio sulla rete Internet pubblica.

I punti di ancoraggio privati consentono di utilizzare gli intervalli di indirizzi IP per la connettività di ancoraggio. I punti di ancoraggio privati vengono creati in una sottorete privata all'interno di un VPC dedicato utilizzando indirizzi IP assegnati dal cliente. Il VPC viene creato nella risorsa Account AWS che possiede la risorsa Outpost e sei responsabile di garantire che il VPC sia disponibile e configurato correttamente (non eliminarlo!). È necessario accedere ai punti di ancoraggio privati utilizzando le VIF private Direct Connect.

È necessario predisporre percorsi di rete ridondanti tra Outpost e i punti di ancoraggio della Regione con connessioni che terminino su dispositivi separati in più di una posizione. Il routing dinamico deve essere configurato per reindirizzare automaticamente il traffico verso percorsi alternativi in caso di guasto delle connessioni o dei dispositivi di rete. È necessario fornire una capacità di rete sufficiente per garantire che l'errore di un percorso WAN non sovraccarichi i percorsi rimanenti.

Il diagramma seguente mostra tre Outposts con percorsi di rete ridondanti verso le rispettive AZ di ancoraggio, AWS Direct Connect utilizzando anche la connettività Internet pubblica. L'Outpost A e l'Outpost B sono ancorati a diverse zone di disponibilità nella stessa regione. L'Outpost A si collega ai punti di ancoraggio privati nella zona AZ 1 della regione 1. L'avamposto B si collega ai punti di ancoraggio pubblici nella zona AZ 2 della regione 1. L'avamposto C si collega agli ancoraggi pubblici in AZ 1 della regione 2.



Connettività di ancoraggio ad alta disponibilità AWS Direct Connect e accesso pubblico a Internet

Outpost A dispone di tre percorsi di rete ridondanti per raggiungere il punto di ancoraggio privato. Sono disponibili due percorsi tramite circuiti Direct Connect ridondanti in un'unica posizione Direct Connect. Il terzo percorso è disponibile tramite un circuito Direct Connect in una seconda posizione Direct Connect. Questo design mantiene il traffico dei collegamenti di servizio di Outpost A sulle reti

private e fornisce una ridondanza dei percorsi che consente il guasto di uno qualsiasi dei circuiti Direct Connect o il guasto di un'intera posizione Direct Connect.

Outpost B dispone di quattro percorsi di rete ridondanti per raggiungere il punto di ancoraggio pubblico. Sono disponibili tre percorsi tramite VIF pubblici forniti sui circuiti e sulle postazioni Direct Connect utilizzati da Outpost A. Il quarto percorso è disponibile tramite la WAN del cliente e la rete Internet pubblica. Il traffico dei link di servizio di Outpost B può essere instradato su qualsiasi percorso disponibile in grado di raggiungere con successo i punti di ancoraggio sulla rete Internet pubblica. L'utilizzo dei percorsi Direct Connect può fornire una latenza più costante e una maggiore disponibilità di larghezza di banda, mentre il percorso Internet pubblico può essere utilizzato per scenari di Disaster Recovery (DR) o di aumento della larghezza di banda.

Outpost C dispone di due percorsi di rete ridondanti per raggiungere il punto di ancoraggio pubblico. Outpost C è distribuito in un data center diverso da Outposts A e B. Il data center di Outpost C non dispone di circuiti dedicati che si collegano alla WAN del cliente. Il data center dispone invece di connessioni Internet ridondanti fornite da due diversi provider di servizi Internet (ISP). Il traffico di collegamento ai servizi di Outpost C può essere instradato su una delle reti ISP per raggiungere i punti di ancoraggio sulla rete Internet pubblica. Questo design consente la flessibilità necessaria per instradare il traffico dei collegamenti di servizio su qualsiasi connessione Internet pubblica disponibile. Tuttavia, il end-to-end percorso dipende dalle reti pubbliche di terze parti in cui la disponibilità della larghezza di banda e la latenza di rete variano.

Il percorso di rete tra un Outpost e i relativi punti di ancoraggio dei collegamenti di servizio deve soddisfare le seguenti specifiche di larghezza di banda:

- 500 Mbps - 1 Gbps di larghezza di banda disponibile per rack Outpost (ad esempio, 3 rack: larghezza di banda disponibile da 1,5 a 3 Gbps)

Pratiche consigliate per una connettività di ancoraggio ad alta disponibilità:

- Fornisci percorsi di rete ridondanti tra ogni avamposto e i relativi punti di ancoraggio nella regione.
- Usa i percorsi Direct Connect (DX) per controllare la latenza e la disponibilità della larghezza di banda.
- Assicurati che le porte TCP e UDP 443 siano aperte (in uscita) dai blocchi CIDR di Outpost Service Link agli intervalli di indirizzi IP EC2 nella regione principale. Assicurati che le porte siano aperte su tutti i percorsi di rete.
- Assicurati che ogni percorso soddisfi i requisiti di disponibilità e latenza della larghezza di banda.

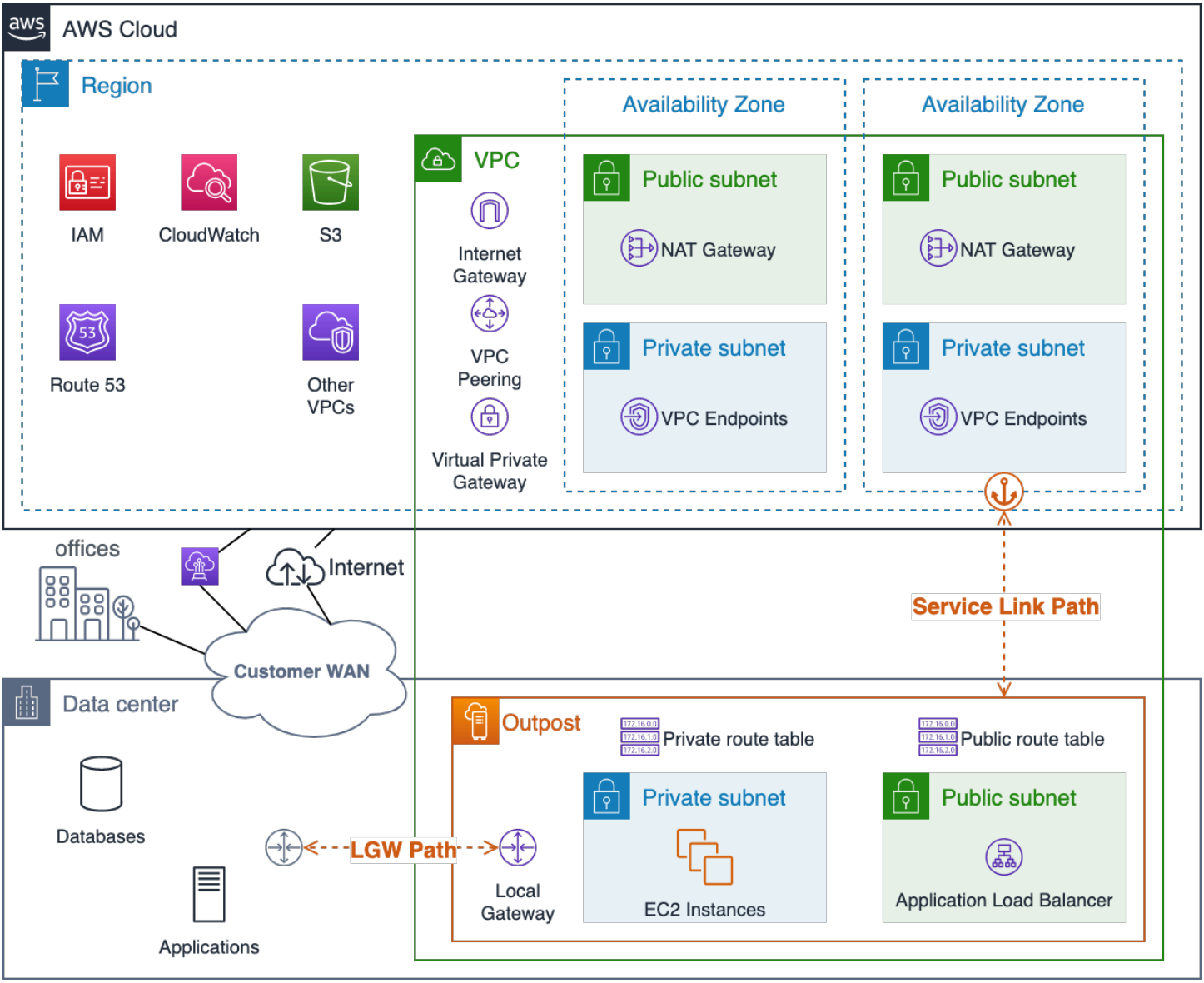
- Utilizza il routing dinamico per automatizzare il reindirizzamento del traffico in caso di guasti di rete.
- Prova a instradare il traffico del service link su ogni percorso di rete pianificato per garantire che il percorso funzioni come previsto.

Routing dell'applicazione/del carico di lavoro

Esistono due percorsi di uscita da Outpost per i carichi di lavoro delle applicazioni:

- Il percorso del collegamento al servizio
- Il percorso del gateway locale (LGW)

È possibile configurare le tabelle di routing delle sottoreti di Outpost per controllare il percorso da seguire per raggiungere le reti di destinazione. Le rotte indirizzate alla LGW indirizzeranno il traffico fuori dal Local Gateway e verso la rete locale. I percorsi indirizzati ai servizi e alle risorse della Regione, come Internet Gateway, NAT Gateway, Virtual Private Gateway e TGW, utilizzeranno service link per raggiungere questi obiettivi. Se disponi di una connessione peering VPC con più VPC sullo stesso Outpost, il traffico tra i VPC rimane sull'Outpost e non utilizza il collegamento di servizio alla Regione. Per informazioni sul peering VPC, consulta Connect VPC utilizzando il peering VPC nella Amazon VPC User Guide.



Visualizzazione del collegamento al servizio Outpost e dei percorsi di rete LGW

Quando si pianifica il routing delle applicazioni, è necessario prestare attenzione a considerare sia il normale funzionamento che il routing limitato e la disponibilità del servizio in caso di guasti di rete. Il percorso Service Link non è disponibile quando un Outpost è disconnesso dalla regione.

È necessario fornire percorsi diversi e configurare il routing dinamico tra Outpost LGW e le applicazioni, i sistemi e gli utenti locali critici. I percorsi di rete ridondanti consentono alla rete di indirizzare il traffico in caso di guasti e garantiscono che le risorse locali siano in grado di comunicare con i carichi di lavoro in esecuzione su Outpost in caso di guasti parziali della rete.

Le configurazioni dei percorsi VPC di Outpost sono statiche. Le tabelle di routing delle subnet vengono configurate tramite AWS Management Console CLI, API e altri strumenti Infrastructure as Code (IaC); tuttavia, non sarà possibile modificare le tabelle di routing della sottorete durante un evento di disconnessione. Dovrai ristabilire la connettività tra Outpost e Region per aggiornare le tabelle di routing. Utilizzate per le normali operazioni gli stessi percorsi che intendete utilizzare durante gli eventi di disconnessione.

Le risorse dell'Outpost possono accedere a Internet tramite il collegamento di servizio e un Internet Gateway (IGW) nella regione o tramite il percorso Local Gateway (LGW). Il routing del traffico Internet sul percorso LGW e sulla rete locale consente di utilizzare i punti di ingresso/uscita Internet esistenti in sede e può fornire latenza inferiore, MTU più elevate e costi di uscita AWS dati ridotti rispetto all'utilizzo del percorso di collegamento del servizio verso un IGW nella regione.

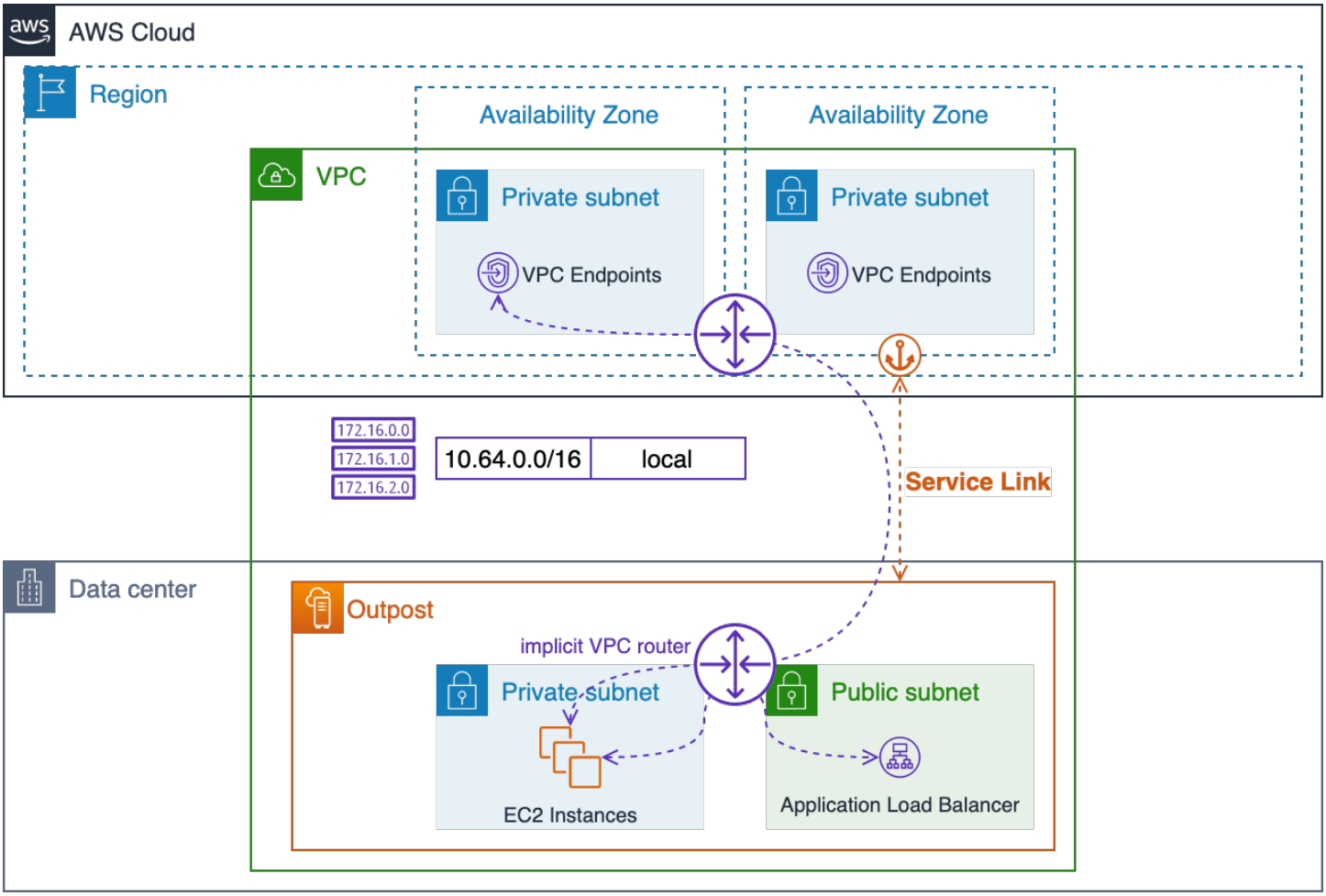
Se l'applicazione deve essere eseguita in locale e deve essere accessibile dalla rete Internet pubblica, è necessario indirizzare il traffico dell'applicazione tramite le connessioni Internet locali verso LGW per raggiungere le risorse su Outpost.

Sebbene sia possibile configurare le sottoreti su un Outpost come le sottoreti pubbliche nella regione, questa può essere una pratica indesiderabile per la maggior parte dei casi d'uso. Il traffico Internet in entrata arriverà attraverso il collegamento del servizio Regione AWS e verrà indirizzato alle risorse in esecuzione su Outpost tramite il collegamento di servizio.

Il traffico di risposta verrà a sua volta instradato tramite il collegamento del servizio e reindirizzato attraverso le connessioni Internet dell' Regione AWS utente. Questo andamento del traffico può aumentare la latenza e comporterà l'addebito di costi per i dati in uscita quando il traffico esce dalla regione per dirigersi verso l'avamposto e quando il traffico di ritorno attraversa la regione ed esce verso Internet. Se l'applicazione può essere eseguita nella regione, quest'ultima è il posto migliore per eseguirla.

Il traffico tra le risorse VPC (nello stesso VPC) seguirà sempre il percorso CIDR VPC locale e verrà instradato tra le sottoreti dai router VPC impliciti.

Ad esempio, il traffico tra un'istanza EC2 in esecuzione su Outpost e un endpoint VPC nella regione verrà sempre instradato tramite il collegamento di servizio.



Routing VPC locale attraverso i router impliciti

Pratiche consigliate per il routing delle applicazioni e dei carichi di lavoro:

- Se possibile, utilizzate il percorso Local Gateway (LGW) anziché il percorso del collegamento al servizio.
- Indirizza il traffico Internet sul percorso LGW.
- Configura le tabelle di routing della sottorete Outpost con un set standard di rotte: verranno utilizzate sia per le normali operazioni che durante gli eventi di disconnessione.
- Fornisci percorsi di rete ridondanti tra Outpost LGW e le risorse critiche delle applicazioni locali. Utilizza il routing dinamico per automatizzare il reindirizzamento del traffico in caso di guasti di rete locali.

Calcolo

Mentre la capacità in ingresso di Amazon EC2 Regioni AWS è apparentemente infinita, la capacità di Outposts è limitata. Sei responsabile della pianificazione e della gestione della capacità di calcolo delle tue distribuzioni Outposts.

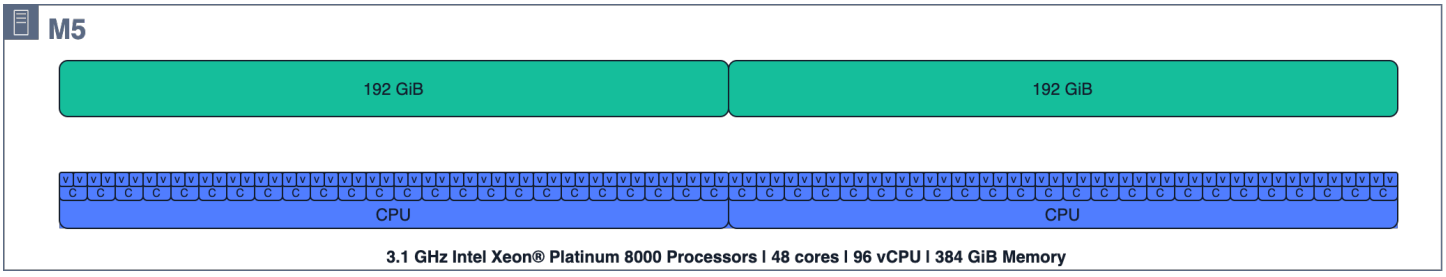
Argomenti

Pianificazione della capacità

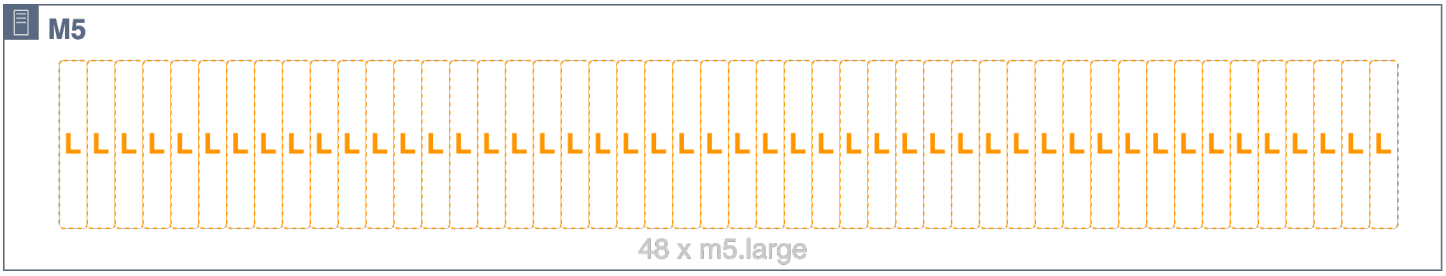
Sebbene la capacità in ingresso di Amazon EC2 Regioni AWS sia apparentemente infinita, la capacità di Outposts è limitata, limitata dal volume totale di capacità di elaborazione ordinata. Sei responsabile della pianificazione e della gestione della capacità di calcolo delle tue distribuzioni Outposts. È necessario ordinare una capacità di elaborazione sufficiente per supportare un modello di disponibilità N+M, dove N è il numero richiesto di server e M è il numero di server di riserva predisposti per far fronte ai guasti del server. N+1 e N+2 sono i livelli di disponibilità più comuni.

Ogni server (`C5`, `M5R5`, ecc.) supporta una singola famiglia di istanze EC2. Prima di poter avviare le istanze sui server di calcolo EC2, devi fornire layout di slot che specifichino le dimensioni delle istanze EC2 che desideri che ciascun server fornisca. AWS configura ogni server con il layout di slotting richiesto.

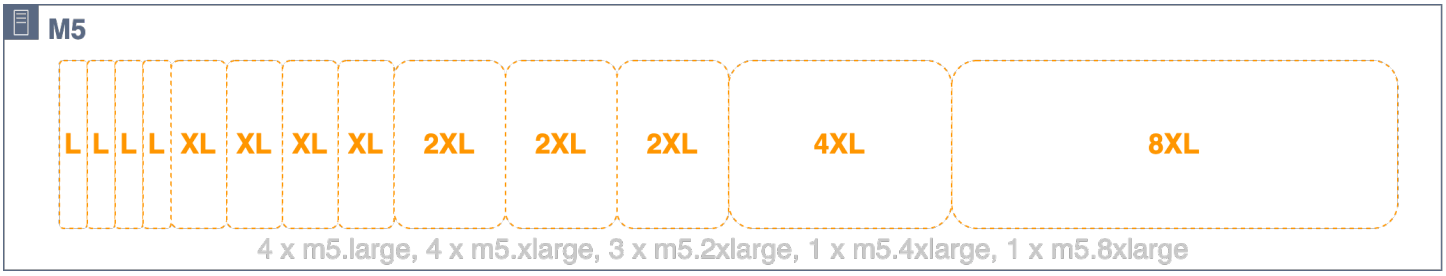
I server possono essere inseriti in modo omogeneo laddove tutti gli slot abbiano la stessa dimensione di istanza (ad esempio, 48 `m5.large` slot) o in modo eterogeneo con una combinazione di tipi di istanze (ad esempio 4, 4`m5.large`, 3, 1 e 1 `m5.4xlargem5.8xlarge`). Per visualizzare queste configurazioni di slotting `m5.xlargem5.2xlarge`, vedere le tre figure successive.



*`m5.24xlarge`*risorse di elaborazione del server



*m5.24xlarge*server inserito in modo omogeneo in 48 slot *m5.large*



*m5.24xlarge*server suddiviso in modo eterogeneo in 4, 4, 3, 1 e 1 slot *m5.large* *m5.xlarge* *m5.2xlarge* *m5.4xlarge* *m5.8xlarge*

Non è necessario sfruttare l'intera capacità del server. È possibile aggiungere slot a un server con capacità non allocata disponibile. È possibile modificare un layout di slot aprendo un ticket di assistenza. Enterprise Support potrebbe richiedere la chiusura o il riavvio di alcune istanze per completare una richiesta di riassegnazione se il nuovo layout di slot non può essere applicato mentre determinati slot sono occupati da istanze in esecuzione.

Tutti i server contribuiscono con gli slot assegnati ai pool di capacità EC2 di Outpost e tutti gli slot di un determinato tipo e dimensione di istanza vengono gestiti come un unico pool di capacità EC2. Ad esempio, il precedente server con slot eterogenei e dotato di`m5.large`,, e slot forniva questi `m5.8xlarge` slot a cinque pool di capacità EC2 `m5.xlarge` `m5.2xlarge`m`5.4xlarge`, un pool per ogni tipo e dimensione di istanza.

È importante considerare lo slotting dei server e i pool di capacità EC2 quando si pianifica la capacità di riserva per la disponibilità dei server N+M. AWS rileva quando un server si guasta o è danneggiato e pianifica una visita al sito per sostituire il server guasto. È necessario progettare i pool di capacità EC2 in modo da tollerare il fallimento di almeno un server di ogni famiglia di istanze (N+1) in un Outpost. Con questo livello minimo di disponibilità del server, quando un server si guasta o deve essere messo fuori servizio, è possibile riavviare le istanze guaste o danneggiate negli slot di riserva dei server rimanenti della stessa famiglia.

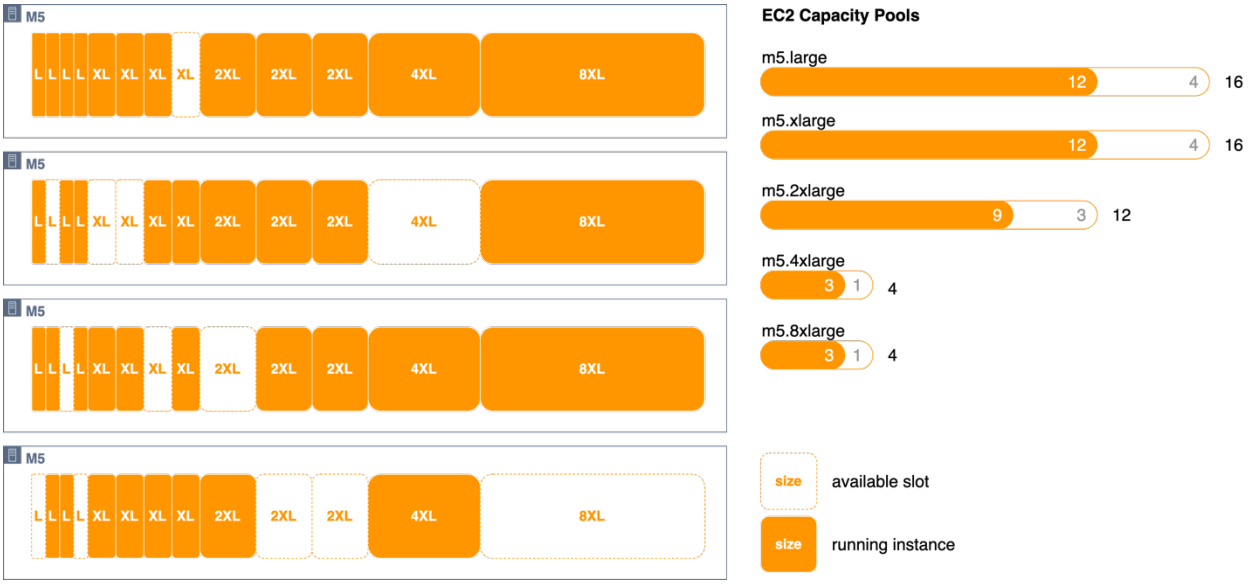
La pianificazione della disponibilità di N+M è semplice quando si dispone di server con slot omogenei o gruppi di server con slot eterogenei con layout di slot identici. È sufficiente calcolare il numero di server (N) necessari per eseguire tutti i carichi di lavoro e quindi aggiungere (M) server aggiuntivi per soddisfare i requisiti di disponibilità dei server durante gli eventi di guasto e manutenzione.

Le seguenti configurazioni di slot non sono utilizzabili a causa dei limiti NUMA:

- 3 `m5.8xlarge`
- 1 `m5.16xlarge` e 1 `m5.8xlarge`

Consultate il vostro Account AWS team per convalidare la configurazione di AWS Outposts rack slotting pianificata.

Nella figura seguente, quattro `m5.24xlarge` server sono suddivisi in modo eterogeneo con un layout di slotting identico. I quattro server creano cinque pool di capacità EC2. Ogni pool funziona al massimo utilizzo (75%) per mantenere la disponibilità N+1 per le istanze in esecuzione su questi quattro server. In caso di guasto di un server, c'è spazio sufficiente per riavviare le istanze fallite sui server rimanenti.



Visualizzazione degli slot dei server EC2, delle istanze in esecuzione e dei pool di slot

Per layout di slot più complessi, in cui i server non hanno slot identici, sarà necessario calcolare la disponibilità N+M per ogni pool di capacità EC2. Puoi utilizzare la seguente formula per calcolare quanti server (che forniscono slot a un determinato pool di capacità EC2) possono fallire e consentire comunque ai server rimanenti di trasportare le istanze in esecuzione:

$$M = \left\lceil \frac{poolSlots_{available}}{serverSlots_{max}} \right\rceil$$

Dove:

- PoolSlots $_{available}$ è il numero di slot disponibili nel pool di capacità EC2 specificato (numero totale di slot nel pool meno il numero di istanze in esecuzione)
- ServerSlots $_{max}$ è il numero massimo di slot forniti da qualsiasi server al pool di capacità EC2 specificato
- M è il numero di server che possono fallire e consentire comunque ai server rimanenti di trasportare le istanze in esecuzione

Esempio: un Outpost ha tre server che forniscono slot a un pool di `m5.2xlarge` capacità. Il primo fornisce 4 slot, il secondo fornisce 3 slot e il terzo server offre 2 slot. Il pool di `m5.2xlarge` istanze di Outpost ha una capacità totale di 9 slot (4 + 3 + 2). Outpost ha 4 istanze in esecuzione`m5.2xlarge`. Quanti server potrebbero fallire e consentire comunque ai server rimanenti di trasportare le istanze in esecuzione?

$$poolSlots_{available} = total\ capacity - running\ instances = 9 - 4 = 5$$

$$serverSlots_{max} = \max([4, 3, 2]) = 4$$

$$M = \left\lceil \frac{poolSlots_{available}}{serverSlots_{max}} \right\rceil = \left\lceil \frac{5}{4} \right\rceil = [1.25] = 1$$

Risposta: È possibile perdere uno qualsiasi dei server e conservare le istanze in esecuzione sui server rimanenti.

Pratiche consigliate per la pianificazione della capacità di calcolo:

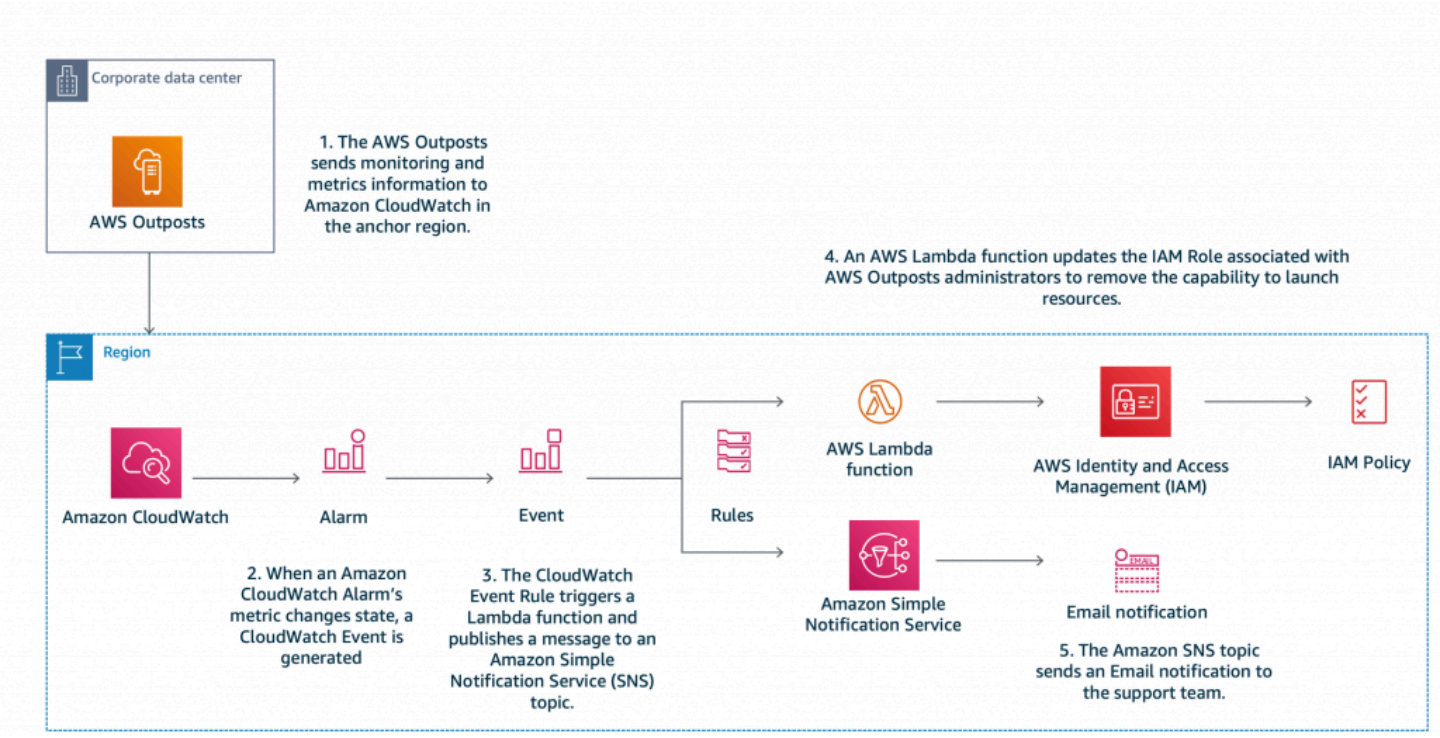
- Dimensiona la tua capacità di elaborazione per fornire ridondanza N+M per ogni pool di capacità EC2 su un Outpost.

- Implementa server N+M per server con slot eterogenei omogenei o identici.
- Calcola la disponibilità N+M per ogni pool di capacità EC2 e assicurati che ogni pool soddisfi i tuoi requisiti di disponibilità.

Gestione della capacità

Puoi monitorare l'utilizzo del pool di istanze Outpost EC2 all'interno e AWS Management Console tramite i parametri di Amazon. CloudWatch Contatta Enterprise Support per recuperare o modificare i layout di slot per i tuoi Outposts.

Utilizzi gli stessi meccanismi di ripristino automatico dell'istanza e di EC2 Auto Scaling per ripristinare o sostituire le istanze influenzate da guasti del server ed eventi di manutenzione. È necessario monitorare e gestire la capacità di Outpost per garantire che sia sempre disponibile una capacità di riserva sufficiente per far fronte ai guasti del server. Il post Managing your AWS Outposts capacity using Amazon CloudWatch and AWS Lambda blog fornisce un tutorial pratico che mostra come combinare AWS CloudWatch e gestire la capacità di Outpost AWS Lambda per mantenere la disponibilità delle istanze.



Gestione AWS Outposts della capacità con Amazon CloudWatch e AWS Lambda

Pratiche consigliate per la gestione della capacità di calcolo:

- Configura le tue istanze EC2 nei gruppi Auto Scaling o utilizza il ripristino automatico delle istanze per riavviare le istanze non riuscite.
- Automatizza il monitoraggio della capacità per le tue implementazioni Outpost e configura le notifiche e (facoltativamente) le risposte automatiche per gli allarmi di capacità.

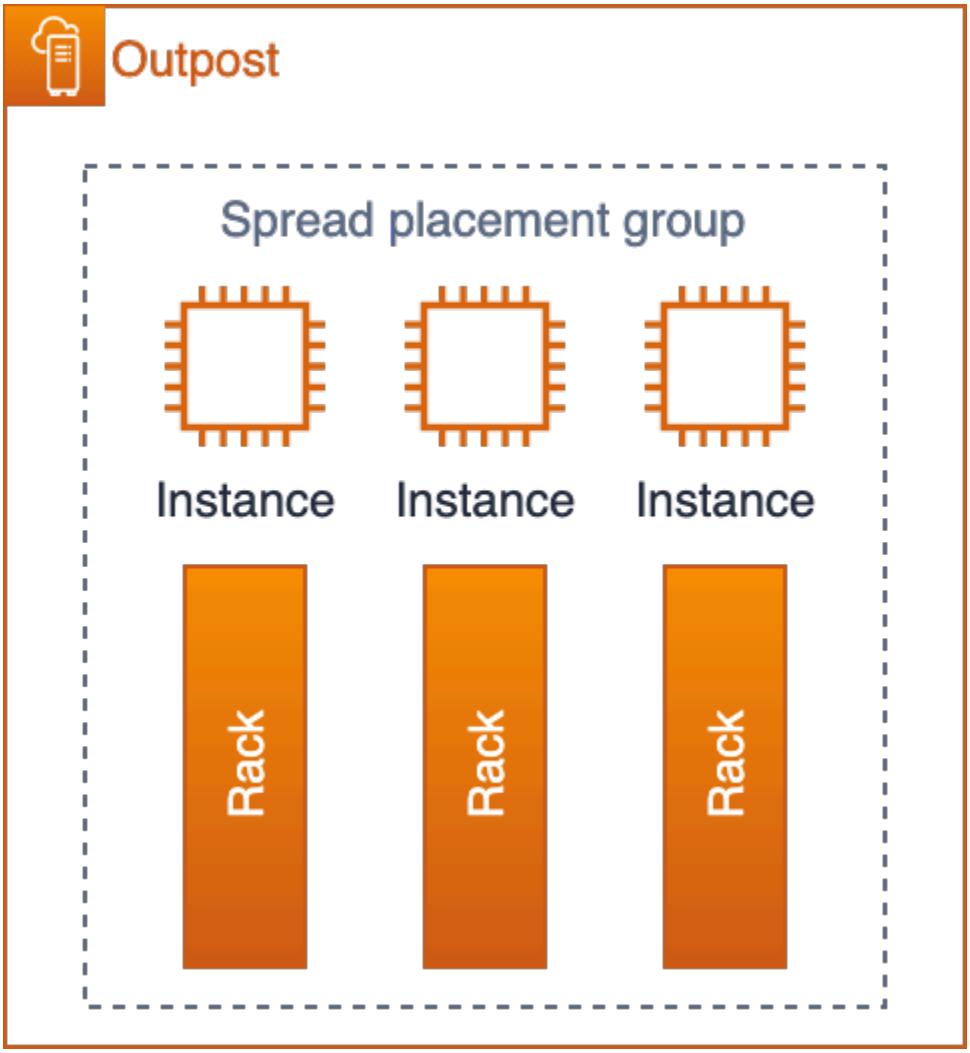
Posizionamento delle varianti

Gli Outposts hanno un numero finito di server di elaborazione. Se l'applicazione distribuisce più istanze correlate su Outposts, senza configurazioni aggiuntive, le istanze possono essere distribuite sullo stesso server o su server nello stesso rack. Oggi, è possibile utilizzare tre meccanismi per distribuire le istanze e ridurre il rischio di eseguire istanze correlate sulla stessa infrastruttura:

Implementazione Multi-Outpost: analogamente a una strategia Multi-AZ nella regione, puoi distribuire Outposts in data center separati e distribuire risorse applicative in Outposts specifici. Ciò consente di eseguire le istanze sull'Outpost desiderato (un set logico di rack). È possibile utilizzare una strategia Multi-Outpost per proteggere dalle modalità di guasto dei rack e dei data center e, se gli Outposts sono ancorati a AZ o regioni separate, può anche fornire protezione contro le modalità di errore AZ o Region. Per ulteriori informazioni sulle architetture Multi-Outpost, consulta la sezione Larger Failure Modes.

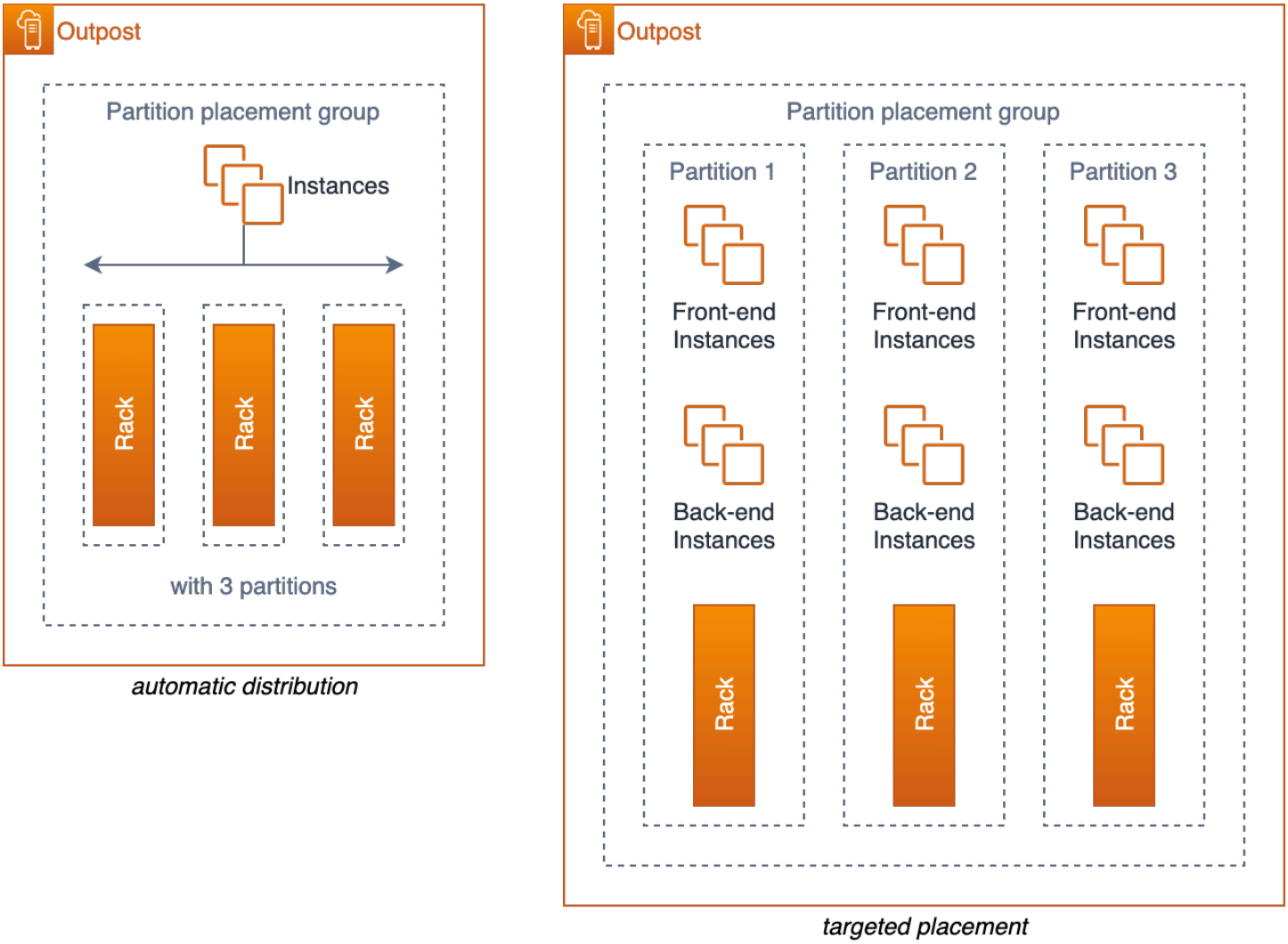
Gruppi di posizionamento di Amazon EC2 su Outposts (posizionamento di istanze multi-rack Single-Outpost): consentono di utilizzare le strategie di cluster, spread e partizione per influenzare il posizionamento. Le strategie di distribuzione e posizionamento delle partizioni consentono di distribuire le istanze tra i rack in un Outpost multi-rack.

Lo spread placement group offre un modo semplice per distribuire le singole istanze tra i rack per ridurre il rischio di guasti correlati. Puoi distribuire nel gruppo solo tante istanze quanti sono i rack presenti in Outpost.



Gruppo di posizionamento distribuito EC2 su un Outpost con tre rack

Puoi anche distribuire le istanze su più rack con gruppi di posizionamento delle partizioni. Utilizza la distribuzione automatica per distribuire le istanze tra le partizioni del gruppo o distribuisci le istanze su partizioni di destinazione selezionate. La distribuzione di istanze su partizioni di destinazione consente di distribuire risorse selezionate sullo stesso rack distribuendo altre risorse tra i rack. Ad esempio, se si dispone di un Outpost logico con tre rack, la creazione di un gruppo di posizionamento delle partizioni con tre partizioni consente di distribuire le risorse tra i rack.



Posizionamento dei gruppi di partizione EC2 su un Outpost con tre rack

Slotting creativo sul server: se disponi di un Outpost a rack singolo o se il servizio che utilizzi su Outposts non supporta i gruppi di posizionamento, potresti essere in grado di utilizzare lo slotting creativo per evitare che le istanze vengano distribuite sullo stesso server fisico. Se le istanze correlate hanno le stesse dimensioni delle istanze EC2, potresti essere in grado di inserire i tuoi server in uno slot per limitare il numero di slot di quella dimensione configurati su ciascun server, distribuendo gli slot tra i server. Lo slotting del server limiterà il numero di istanze (di quelle dimensioni) che possono essere eseguite su un singolo server.

Ad esempio, si consideri il layout di slotting mostrato in precedenza nella Figura 13. Se l'applicazione avesse bisogno di distribuire tre `m5.4xlarge` istanze sull'Outpost configurato con questo layout di slotting, EC2 collocherebbe ogni istanza su un server separato e non ci sarebbe alcuna possibilità

che queste istanze possano essere eseguite sullo stesso server, a condizione che la configurazione di slotting non cambi per aprire slot aggiuntivi sui server. `m5.4xlarge`

Pratiche consigliate per il posizionamento delle istanze di calcolo:

- Usa i gruppi di posizionamento di Amazon EC2 su Outposts per controllare il posizionamento delle istanze tra i rack all'interno di un singolo Outpost.
- Invece di ordinare un Outpost con un unico rack Outpost medio o grande, prendi in considerazione la possibilità di suddividere la capacità in due rack piccoli o medi per sfruttare la capacità dei gruppi di collocamento EC2 di distribuire le istanze tra i rack.

Storage

Il servizio AWS Outposts rack offre tre tipi di storage:

- Archiviazione delle istanze su tipi di istanze EC2 supportati
- Volumi gp2 Amazon Elastic Block Store (EBS) per lo storage a blocchi persistente
- Amazon Simple Storage Service on Outposts (S3 on Outposts) per lo storage locale di oggetti

Lo storage delle istanze viene fornito sui server supportati (`C5d`,, `M5d` `R5dG4dn`, e). `I3en` Proprio come nella regione, i dati in un instance store persistono solo per la durata (in esecuzione) dell'istanza.

I volumi EBS di Outposts e lo storage di oggetti S3 on Outposts sono forniti come parte dei servizi gestiti su rack. AWS Outposts I clienti sono responsabili della gestione della capacità dei pool di storage Outpost. I clienti specificano i requisiti di archiviazione per lo storage EBS e S3 quando ordinano un Outpost. AWS configura Outpost con il numero di server di storage necessari per fornire la capacità di archiviazione richiesta. AWS è responsabile della disponibilità dei servizi di storage EBS e S3 on Outposts. Viene fornito un numero sufficiente di server di storage per fornire servizi di storage ad alta disponibilità all'Outpost. La perdita di un singolo server di storage non deve interrompere i servizi né comportare la perdita di dati.

Puoi utilizzare le CloudWatch metriche AWS Management Console and per monitorare l'utilizzo della capacità di Outpost EBS e S3 on Outposts.

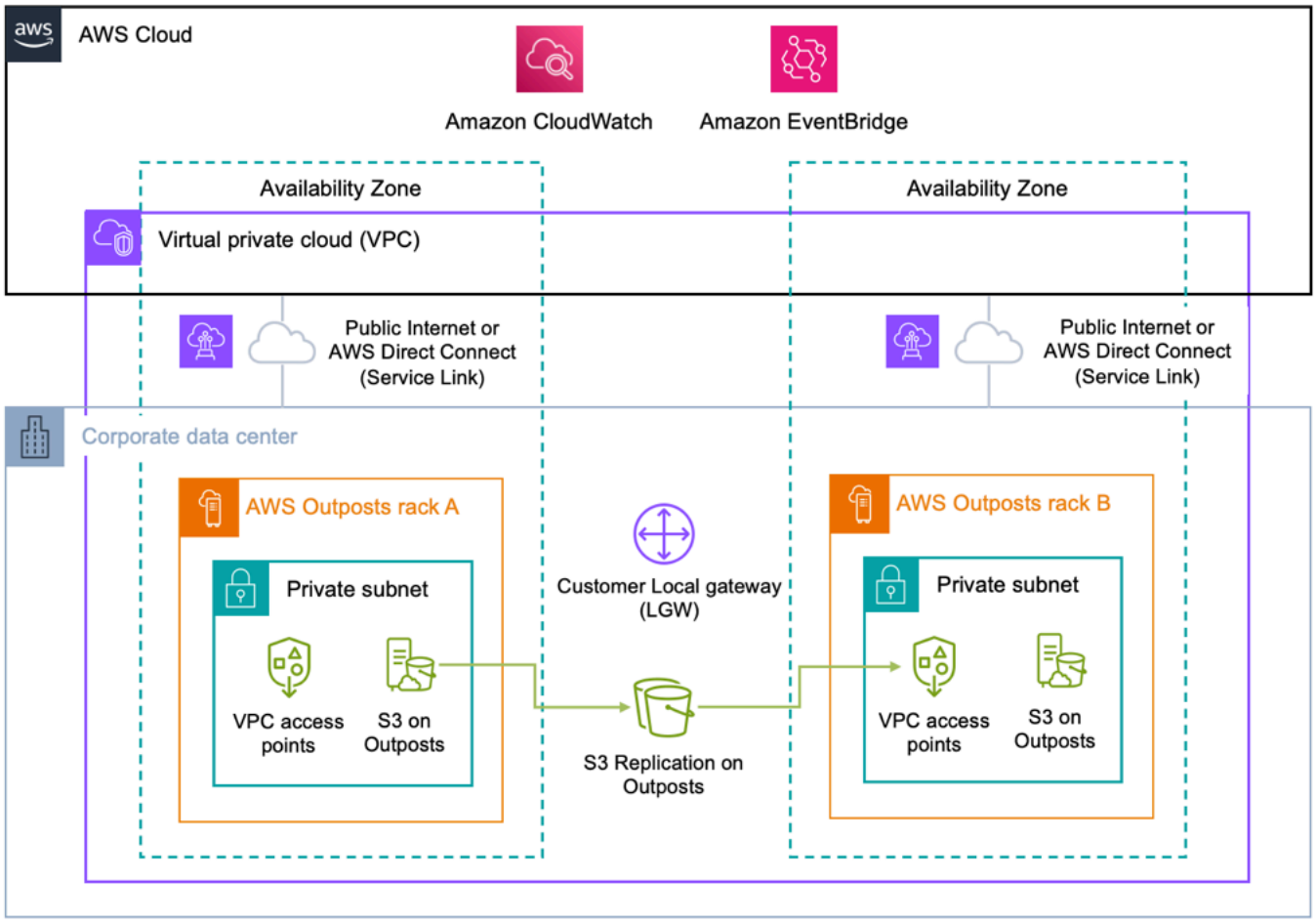
Protezione dei dati

Per EBS Volumes: AWS Outposts rack supporta le istantanee dei volumi EBS per fornire un meccanismo di protezione dei dati semplice e sicuro per proteggere i dati di storage a blocchi. Le istantanee sono backup point-in-time incrementali dei volumi EBS. Per impostazione predefinita, le istantanee dei volumi Amazon EBS su Outpost vengono archiviate su Amazon S3 nella regione. Se i tuoi Outposts sono stati configurati con la capacità di S3 on Outposts, puoi utilizzare EBS Local Snapshots on Outposts per archiviare le istantanee localmente sul tuo Outpost utilizzando lo storage S3 on Outposts.

Per i bucket S3 on Outposts (casi d'uso relativi alla residenza dei dati):

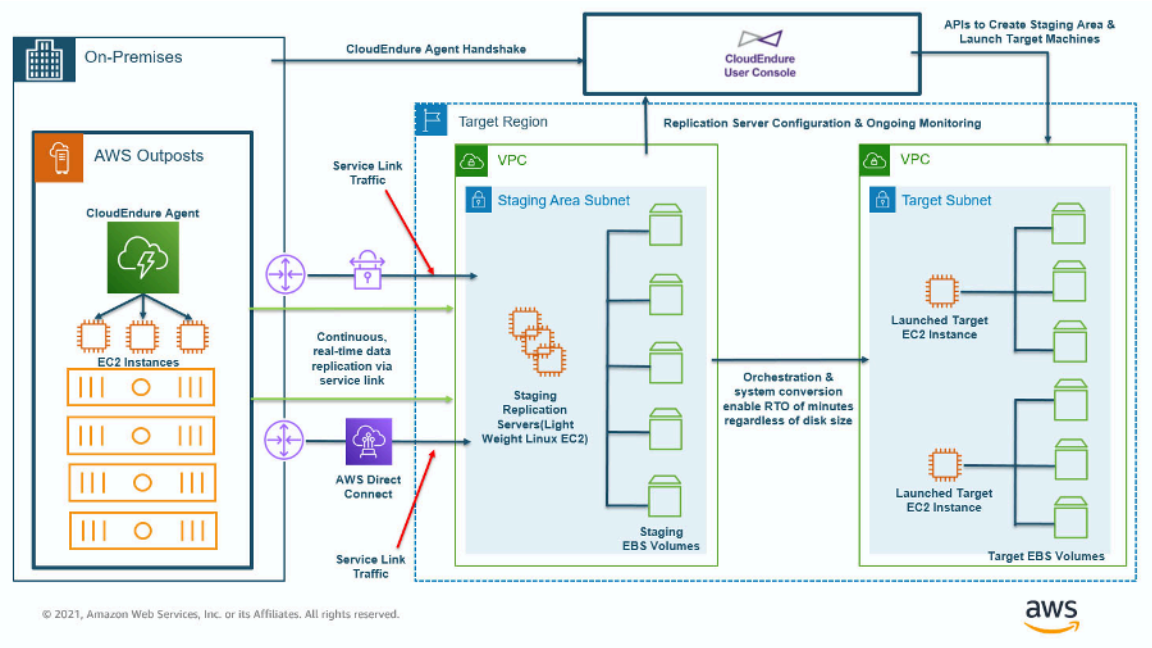
- Puoi usare S3 Versioning su Outposts, per salvare tutte le modifiche e la cronologia degli oggetti. Se abilitato, il controllo delle versioni S3 conserva più copie distinte di un oggetto nello stesso bucket. Puoi utilizzare il controllo delle versioni S3 per conservare, recuperare e ripristinare qualsiasi versione di ogni oggetto archiviato nei bucket Outposts. Il controllo delle versioni S3 ti consente di eseguire il ripristino a seguito di errori dell'applicazione e operazioni non intenzionali degli utenti.
- Puoi utilizzare S3 Replication on Outposts per creare e configurare regole di replica per replicare automaticamente gli oggetti S3 su un altro Outpost o su un altro bucket sullo stesso Outpost. Durante la replica, gli oggetti S3 on Outposts vengono inviati tramite il gateway locale (LGW) del cliente e gli oggetti non tornano indietro al. Regione AWS S3 Replication on Outposts offre un modo semplice e flessibile per replicare automaticamente i dati all'interno di uno specifico perimetro di dati per soddisfare i requisiti di ridondanza e conformità dei dati.

  S3 Replication on Outposts fornisce anche metriche e notifiche dettagliate per monitorare lo stato della replica degli oggetti. Puoi monitorare l'avanzamento della replica monitorando i byte in sospeso, le operazioni in sospeso e la latenza di replica tra i bucket Outposts di origine e di destinazione utilizzando Amazon. CloudWatch Puoi anche configurare EventBridge le regole di Amazon per ricevere eventi di errore di replica per diagnosticare e correggere rapidamente i problemi di configurazione.



Per i bucket S3 on Outposts (casi d'uso non legati alla residenza dei dati) Regioni AWS a: Puoi utilizzarli per AWS DataSyncautomatizzare i trasferimenti di dati di S3 on Outposts tra Outpost e la regione. DataSync ti consente di scegliere cosa trasferire, quando trasferire e quanta larghezza di banda utilizzare. Il backup dei bucket S3 on premise Outposts su bucket S3 in the Regione AWS consente di sfruttare il 99,99999% (11 9) di durabilità dei dati e livelli di storage aggiuntivi (Standard, Infrequent Access e Glacier) per l'ottimizzazione dei costi disponibili con il servizio S3 regionale.

Replica delle istanze: è possibile utilizzarla CloudEndureper replicare singole istanze dai sistemi locali a un avamposto, da un avamposto alla regione, dalla regione a un avamposto o da un avamposto all'altro. Il post sul CloudEndure blog Architecting for DR on AWS Outposts with descrive ciascuno di questi scenari e come progettare una soluzione con. CloudEndure



Disaster recovery (DR) da un avamposto alla regione

L'utilizzo del AWS Outposts rack come CloudEndure destinazione (destinazione di replica) richiede lo storage S3 on Outposts.

Pratiche consigliate per la protezione dei dati:

- Usa gli snapshot EBS per creare point-in-time backup di volumi di storage a blocchi su Amazon S3 nella regione o S3 su Outposts.
- Usa il controllo delle versioni degli oggetti di S3 on Outposts per mantenere più versioni e cronologia dei tuoi oggetti.
- Usa S3 Replication on Outposts per replicare automaticamente i dati degli oggetti su un altro Outpost.
- Per casi d'uso diversi dalla residenza dei dati, utilizza AWS DataSync per eseguire il backup degli oggetti archiviati in S3 su Outpost su Amazon S3 nella regione.
- CloudEndure Da utilizzare per replicare le istanze tra sistemi locali, Outposts logici e la regione.

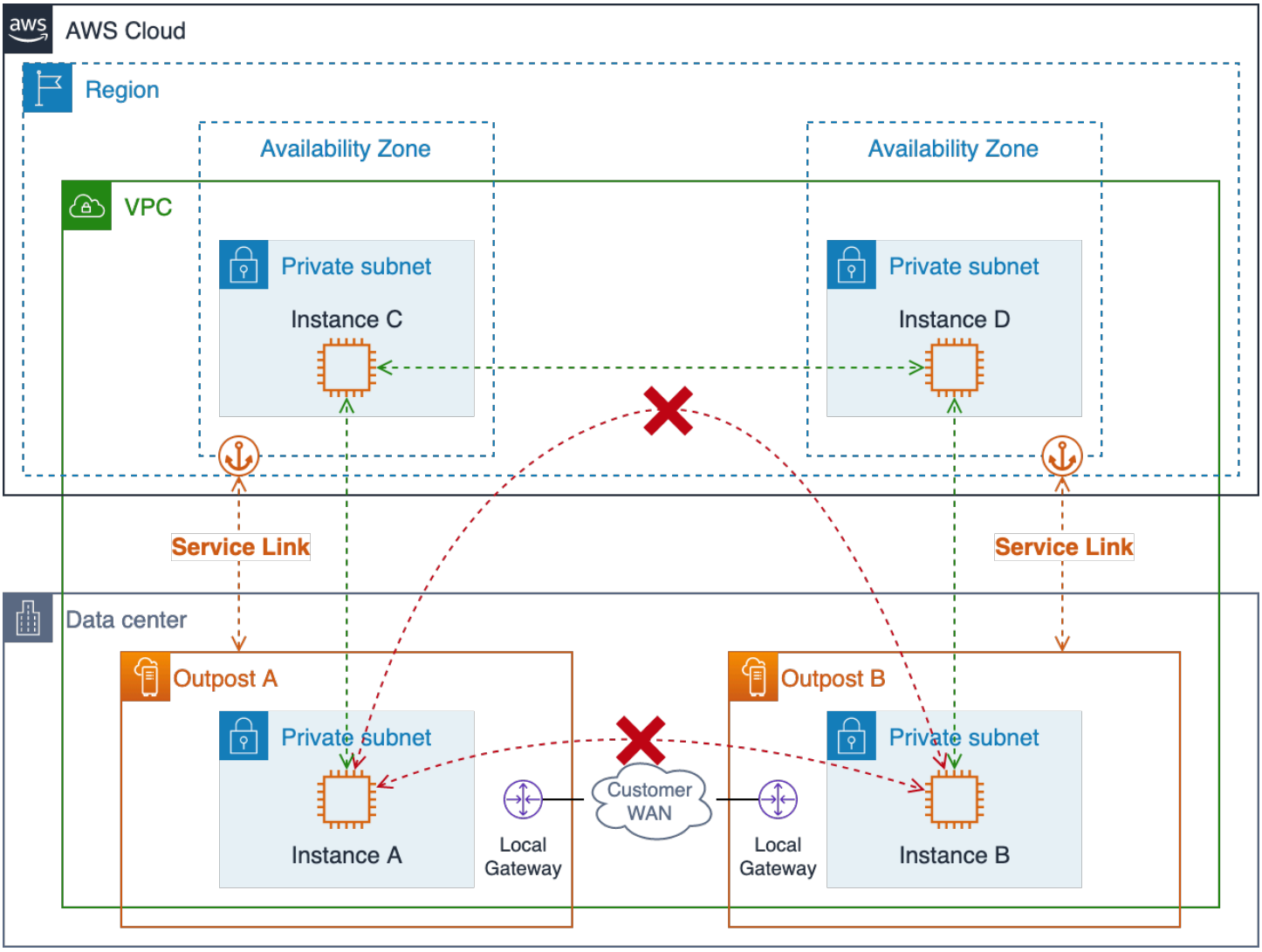
Modalità di errore più ampie

Per progettare architetture HA per mitigare modalità di guasto più ampie come guasti in rack, data center, Availability Zone (AZ) o Region, è necessario implementare più Outposts con una capacità di

infrastruttura sufficiente in data center separati con alimentazione e connettività WAN indipendenti. Gli Outposts vengono ancorati a diverse zone di disponibilità (AZ) all'interno di Regione AWS una o più regioni. È inoltre necessario fornire una site-to-site connettività resiliente e sufficiente tra le sedi per supportare la replica sincrona o asincrona dei dati e il reindirizzamento del traffico dei carichi di lavoro. A seconda dell'architettura dell'applicazione, puoi utilizzare i servizi DNS Amazon Route 53 disponibili a livello globale e i servizi Elastic Load Balancing disponibili a livello regionale per indirizzare il traffico verso la posizione desiderata e automatizzare il reindirizzamento del traffico verso le località sopravvissute in caso di guasti su larga scala.

Esistono limitazioni di rete di cui è necessario tenere conto durante la progettazione e la distribuzione di carichi di lavoro delle applicazioni su più Outposts. Le risorse su due Outposts separati non possono comunicare tra loro facendo transitare il traffico attraverso la Regione. Le risorse su due Outposts separati distribuiti all'interno dello stesso VPC non possono comunicare tra loro attraverso la rete del cliente. Le risorse su due Outposts separati distribuiti in VPC diversi possono comunicare tra loro attraverso la rete del cliente.

Le due figure seguenti illustrano i percorsi di rete bloccati e riusciti.

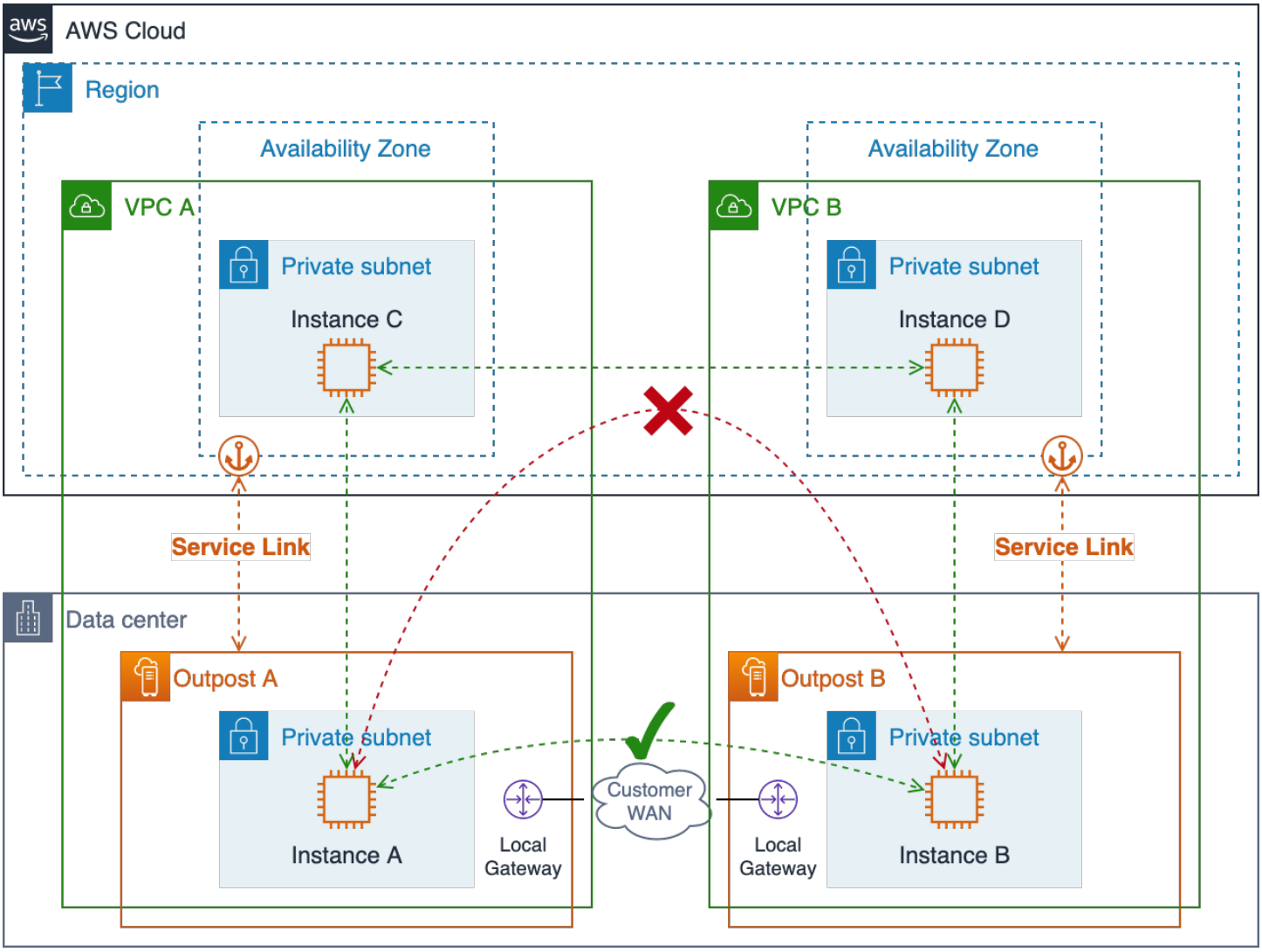


Percorsi di rete VPC singolo con più avamposti

Il traffico da avamposto a avamposto che transita nella regione è bloccato in quanto si tratta di un anti-pattern. Tale traffico comporterebbe costi di uscita in entrambe le direzioni e probabilmente avrebbe una latenza molto più elevata rispetto al semplice instradamento del traffico attraverso la WAN del cliente.

Le risorse su più Outposts nello stesso VPC non possono comunicare tra loro. Il traffico tra Outpost nello stesso VPC seguirà sempre il percorso CIDR VPC locale attraverso la regione in cui verrà bloccato.

È necessario utilizzare VPC separati per distribuire risorse su più Outposts per consentire di instradare il traffico da Outpost a Outpost attraverso le reti locali locali locali e WAN.



Percorsi di rete con più VPC e più avamposti

Pratiche consigliate per la protezione da modalità di guasto più ampie:

- Implementa più Outposts ancorati a più AZ e regioni.
- Usa VPC separati per ogni Outpost in una distribuzione multi-Outpost.

Conclusioni

Con AWS Outposts rack, puoi creare, gestire e scalare applicazioni locali ad alta disponibilità utilizzando AWS strumenti e servizi familiari come Amazon EC2, Amazon EBS, Amazon S3 on Outposts, Amazon ECS, Amazon EKS e Amazon RDS. I carichi di lavoro possono essere eseguiti localmente, servire i clienti, accedere ad applicazioni e sistemi nelle reti locali e accedere al set completo di servizi di. Regione AWS Il rack Outposts è ideale per carichi di lavoro che richiedono accesso a bassa latenza ai sistemi locali, elaborazione locale dei dati, residenza dei dati e migrazione di applicazioni con interdipendenze di sistema locali.

Quando si fornisce una distribuzione Outpost con alimentazione, spazio e raffreddamento adeguati e connessioni resilienti a, è possibile creare servizi di data center singoli ad Regione AWS alta disponibilità. Inoltre, per livelli più elevati di disponibilità e resilienza, puoi implementare più Outposts e distribuire le tue applicazioni oltre i confini logici e geografici.

Il rack Outposts elimina il peso indifferenziato della creazione di pool di elaborazione, storage e reti di applicazioni locali e consente di estendere la portata dell'infrastruttura AWS globale ai data center e alle strutture di co-locazione. Ora puoi concentrare il tuo tempo e le tue energie sulla modernizzazione delle applicazioni, sulla semplificazione delle implementazioni delle applicazioni e sull'aumento dell'impatto aziendale dei tuoi servizi IT.

Collaboratori

Hanno collaborato alla stesura del presente documento:

- Mallory Gershenfeld, S3 su Outposts, Amazon Web Services
- Chris Lunsford, Architetto specializzato in soluzioni senior AWS Outposts, Amazon Web Services
- Rohan Mathews, architetto capo AWS Outposts, Amazon Web Services

Cronologia dei documenti

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

Modifica	Descrizione	Data
Aggiornamento secondario	Sono state aggiunte ulteriori indicazioni sullo slotting nella pianificazione della capacità.	9 febbraio 2024
Aggiornamento secondario	Aggiornato per riflettere il lancio di funzionalità dopo la pubblicazione iniziale.	19 luglio 2023
Aggiornamento secondario	Procedure consigliate aggiornate per i collegamenti di rete ad alta disponibilità.	29 giugno 2023
Pubblicazione iniziale	Whitepaper pubblicato per la prima volta.	12 agosto 2021

> **Note**
>
> Per sottoscrivere gli aggiornamenti RSS, è necessario che un plug-in RSS sia abilitato per il browser in uso.

Note

I clienti sono responsabili della propria valutazione indipendente delle informazioni contenute in questo documento. Questo documento: (a) è solo a scopo informativo, (b) rappresenta le offerte e le pratiche attuali di AWS prodotti, che sono soggette a modifiche senza preavviso, e (c) non crea alcun impegno o assicurazione da parte dei suoi affiliati, AWS fornitori o licenzianti. AWS i prodotti o i servizi sono forniti «così come sono» senza garanzie, dichiarazioni o condizioni di alcun tipo, esplicite o implicite. Le responsabilità e le responsabilità dei AWS propri clienti sono regolate da AWS accordi e il presente documento non fa parte di, né modifica, alcun accordo tra AWS e i suoi clienti.

AWS Glossario

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

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