AWS Documento técnico

AWS Outposts Consideraciones de arquitectura y diseño de alta disponibilidad

AWS Outposts Consideraciones de arquitectura y diseño de alta disponibilidad: AWS Documento técnico

Table of Contents

AWS Outposts Consideraciones de arquitectura y diseño de alta disponibilidad

Fecha de publicación: 12 de agosto de 2021 (Historial del documento)

Este documento técnico analiza las consideraciones de arquitectura y las prácticas recomendadas que los administradores de TI y los arquitectos de sistemas pueden aplicar para crear entornos de aplicaciones locales de alta disponibilidad. AWS Outposts

¿Usa Well-Architected?

El marco de AWS Well-Architected le ayuda a entender las ventajas y desventajas de las decisiones que toma al crear sistemas en la nube. Los seis pilares del marco le permitirán aprender las prácticas recomendadas de arquitectura para diseñar y utilizar sistemas fiables, seguros, eficientes, rentables y sostenibles. Mediante AWS Well-Architected Tool, disponible sin costo alguno en la AWS Management Console, puede comparar las cargas de trabajo con estas prácticas recomendadas respondiendo a una serie de preguntas para cada pilar.

Para obtener más orientación experta y prácticas recomendadas para la arquitectura de la nube (implementaciones de arquitectura de referencia, diagramas y documentos técnicos), consulte el Centro de arquitectura de AWS.

Introducción

Este paper está dirigido a los administradores de TI y arquitectos de sistemas que desean implementar, migrar y operar aplicaciones mediante la plataforma en la AWS nube y ejecutar esas aplicaciones en las instalaciones con un AWS Outposts rack, el formato de rack de 42U de AWS Outposts.

Presenta los patrones de arquitectura, los antipatrones y las prácticas recomendadas para crear sistemas de alta disponibilidad que incluyan AWS Outposts racks. Aprenderá a administrar la capacidad de sus AWS Outposts racks y a utilizar los servicios de redes y centros de datos para configurar soluciones de infraestructura de AWS Outposts racks de alta disponibilidad.

AWS Outposts rack es un servicio totalmente gestionado que proporciona un conjunto lógico de capacidades de computación, almacenamiento y redes en la nube. Con los bastidores de Outposts, los clientes pueden utilizar los servicios administrados compatibles de AWS en sus entornos en las

instalaciones, como Amazon Elastic Compute Cloud (Amazon EC2), Amazon Elastic Block Store (Amazon EBS), Amazon S3 en Outposts, Amazon Elastic Kubernetes Service (Amazon EKS), Amazon Elastic Container Service (Amazon ECS), Amazon Relational Database Service (Amazon RDS) y otros servicios de AWS en Outposts. Los servicios de Outposts se prestan en el mismo AWS Nitro System que se utiliza en las Regiones de AWS.

Al aprovechar el AWS Outposts rack, puede crear, administrar y escalar aplicaciones locales de alta disponibilidad mediante herramientas y servicios en AWS la nube que ya conoce. AWS Outposts rack es ideal para cargas de trabajo que requieren acceso de baja latencia a sistemas locales, procesamiento de datos local, residencia de datos y migración de aplicaciones con interdependencias entre sistemas locales.

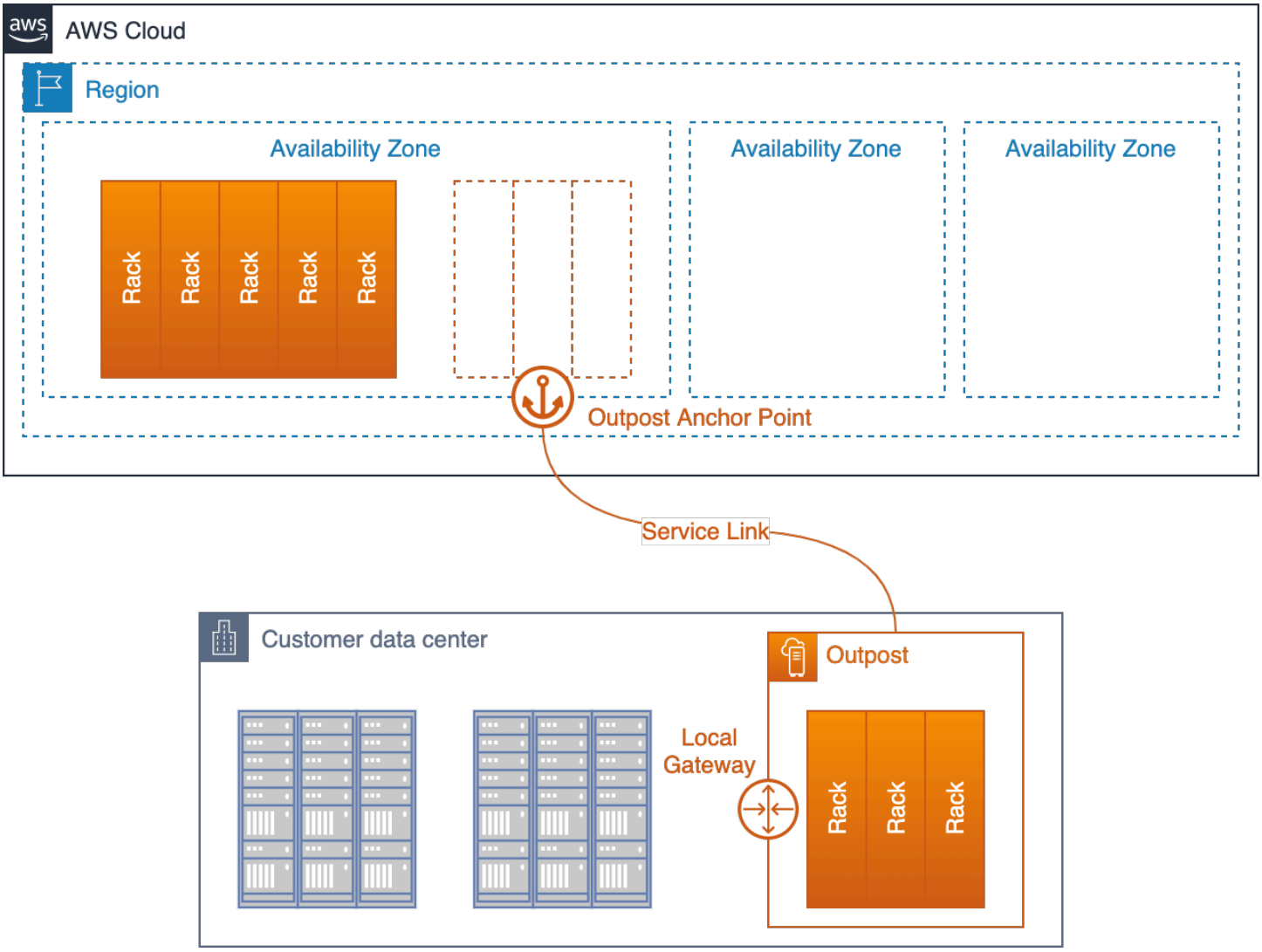
Extender la AWS infraestructura y los servicios a ubicaciones locales

El AWS Outposts servicio ofrece AWS infraestructura y servicios a ubicaciones locales en más de 50 países y territorios, lo que brinda a los clientes la posibilidad de implementar la misma AWS infraestructura, AWS servicios, API y herramientas en prácticamente cualquier centro de datos, espacio de ubicación conjunta o instalación local para lograr una experiencia híbrida verdaderamente coherente. Para entender cómo diseñar con Outposts, debes entender los diferentes niveles que componen la AWS nube.

Una Región de AWS es un área geográfica del mundo. Cada uno Región de AWS es un conjunto de centros de datos que se agrupan de forma lógica en zonas de disponibilidad (AZ). Regiones de AWS proporcionan varias (al menos dos) zonas de disponibilidad aisladas y separadas físicamente que estén conectadas con una conectividad de red redundante, de baja latencia y de alto rendimiento. Cada AZ consta de uno o varios centros de datos físicos.

Un Outpost lógico (en adelante denominado Outpost) es un despliegue de uno o más AWS Outposts racks conectados físicamente y gestionados como una sola entidad. Un Outpost proporciona un conjunto de capacidad AWS informática y de almacenamiento en uno de sus sitios como una extensión privada de una zona de disponibilidad en un. Región de AWS

Quizás el mejor modelo conceptual AWS Outposts sea pensar en desconectar uno o más racks de un centro de datos en una zona de disponibilidad. Región de AWS Los bastidores se implementan desde el centro de datos de la AZ hasta su propio centro de datos. A continuación, tendrá que conectar los bastidores a los puntos de anclaje del centro de datos de la AZ con un cable (muy) largo para que los bastidores sigan funcionando como parte de la Región de AWS. También se conectan a la red local para proporcionar una conectividad de baja latencia entre las redes en las instalaciones y las cargas de trabajo que se ejecutan en dichos bastidores.



Instancia de Outposts implementada en el centro de datos del cliente y que se conecta de nuevo a la AZ de anclaje y la región principal

El Outpost funciona como una extensión de la AZ en la que está anclado. AWS opera, monitorea y administra la AWS Outposts infraestructura como parte de. Región de AWS En lugar de un cable físico muy largo, una instancia de Outposts se conecta a su región principal a través de un conjunto de túneles VPN cifrados denominados enlace de servicio.

El enlace de servicio termina en un conjunto de puntos de anclaje de una zona de disponibilidad (AZ) de la región principal de la instancia de Outposts.

La ubicación donde se almacena su contenido es solo suya. Puede replicar y hacer copias de seguridad de su contenido en esa ubicación Región de AWS o en otras ubicaciones. El contenido no se trasladará ni copiará a otras ubicaciones sin su consentimiento, excepto cuando sea necesario

para cumplir la ley o una orden vinculante de un organismo público. Para obtener más información, consulte Preguntas frecuentes sobre privacidad de datos de AWS.

Las cargas de trabajo que implementa en esos bastidores se ejecutan de forma local. Además, si bien la capacidad de procesamiento y almacenamiento disponible en esos racks es limitada y no permite ejecutar los servicios a escala de nube propios de un rack Región de AWS, los recursos desplegados en el rack (sus instancias y su almacenamiento local) se benefician de la ejecución local mientras el plano de administración sigue funcionando en el mismo. Región de AWS

Para implementar cargas de trabajo en una instancia de Outposts, añada subredes a sus entornos de nube privada virtual (VPC) y especifique una instancia de Outposts como ubicación para las subredes. A continuación, selecciona la subred deseada al implementar AWS los recursos compatibles a través de las AWS Management Console herramientas CLI, API, CDK o infraestructura como código (IaC). Las instancias en las subredes de Outposts se comunican con otras instancias en Outposts o en la región a través de redes VPC.

El enlace de servicio de la instancia de Outposts transporta tanto el tráfico de administración de la instancia en sí como el tráfico de la VPC del cliente (tráfico de la VPC entre las subredes de la instancia de Outposts y las subredes de la región).

Términos importantes:

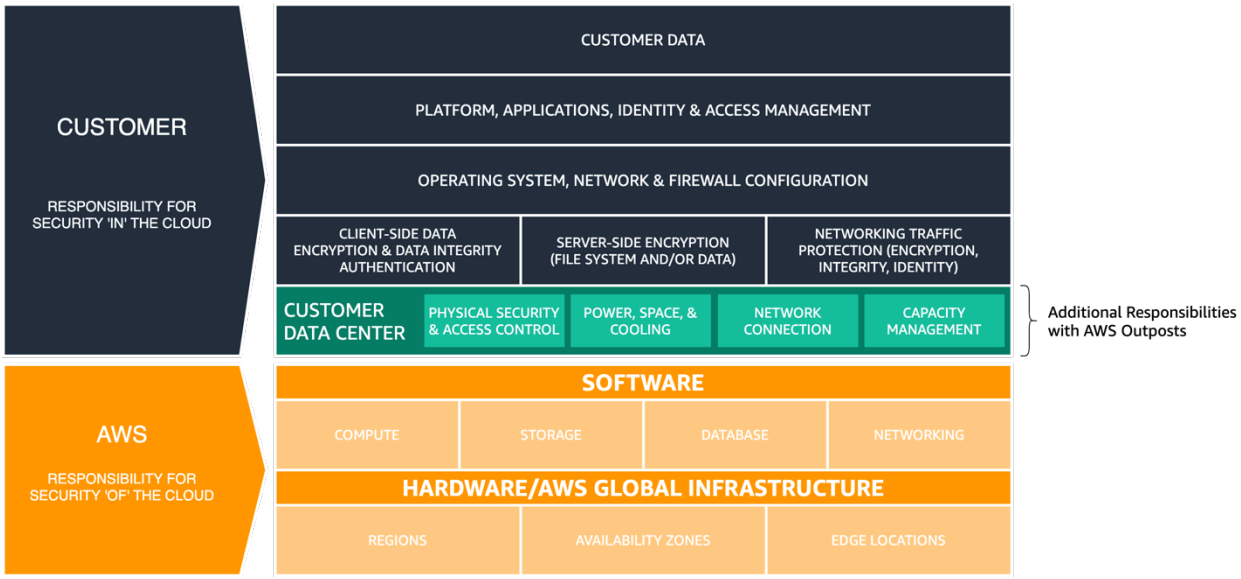
- AWS Outposts— es un servicio totalmente gestionado que ofrece la misma AWS infraestructura, AWS servicios, API y herramientas a prácticamente cualquier centro de datos, espacio compartido o instalación local para ofrecer una experiencia híbrida realmente coherente.
- Outpost: es una implementación de uno o más AWS Outposts racks conectados físicamente que se administra como una entidad lógica única y un conjunto de AWS recursos informáticos, almacenamiento y redes desplegados en las instalaciones de un cliente.
- Región principal: la Región de AWS que proporciona la administración, los servicios de plano de control y los AWS servicios regionales necesarios para un despliegue de Outpost.
- Zona de disponibilidad de anclaje (AZ de anclaje): la zona de disponibilidad de la región principal que aloja los puntos de anclaje de una implementación de Outposts. Una instancia de Outposts funciona como una extensión de su zona de disponibilidad de anclaje.
- Puntos de anclaje: puntos de conexión de la AZ de anclaje que reciben las conexiones de las instancias de Outposts implementadas de forma remota.
- Enlace de servicio: conjunto de túneles VPN cifrados que conectan una instancia de Outposts con su zona de disponibilidad principal de anclaje en su región principal.

- Puerta de enlace local (LGW): un enrutador virtual de interconexión lógica que permite la comunicación entre la instancia de Outposts y la red en las instalaciones del usuario.

Modelo de responsabilidad compartida actualizado

Al implementar la AWS Outposts infraestructura en sus centros de datos o instalaciones compartidas, asume responsabilidades adicionales en el modelo de responsabilidad AWS compartida. Por ejemplo, en la región, AWS ofrece diversas fuentes de alimentación, redes principales redundantes y una conectividad de red de área extendida (WAN) flexible para garantizar la disponibilidad de los servicios en caso de que se produzcan errores en uno o más componentes.

Con Outposts, usted es responsable de garantizar un suministro eléctrico y conectividad de red resilientes a los bastidores de Outposts a fin de satisfacer sus propios requisitos de disponibilidad para las cargas de trabajo que se ejecutan en Outposts.



AWS Modelo de responsabilidad compartida actualizado para AWS Outposts

Con AWS Outposts, usted es responsable de la seguridad física y los controles de acceso del entorno del centro de datos. Debe proporcionar energía, espacio y refrigeración suficientes para mantener la instancia de Outposts y las conexiones de red operativas para volver a conectar la instancia con la región.

Como la capacidad de Outpost es finita y está determinada por el tamaño y la cantidad de racks AWS instalados en su sitio, debe decidir cuánta capacidad de EC2, EBS y S3 en Outpost necesita para ejecutar sus cargas de trabajo iniciales, adaptarse al crecimiento futuro y proporcionar capacidad adicional para mitigar las fallas del servidor y los eventos de mantenimiento.

AWS es responsable de la disponibilidad de la infraestructura de Outposts, incluidas las fuentes de alimentación, los servidores y los equipos de red de los AWS Outposts racks. AWS también administra el hipervisor de virtualización, los sistemas de almacenamiento y los AWS servicios que se ejecutan en Outposts.

En cada bastidor de Outposts se instala un sistema eléctrico central que convierte la corriente alterna en corriente continua y suministra energía a los servidores del bastidor a través de una arquitectura de barras de distribución. Con este tipo de arquitectura, la mitad de las fuentes de alimentación del bastidor pueden fallar sin que ninguno de los servidores interrumpa su funcionamiento.

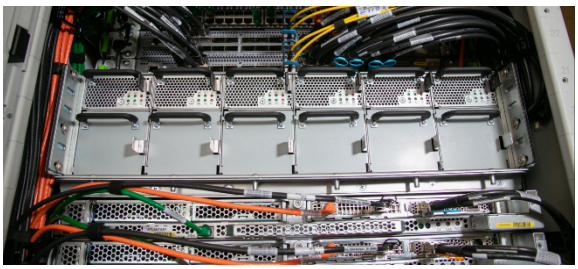


Figura 3: Fuentes de alimentación de AWS Outposts CA a CC y distribución de energía de barra colectora

Los conmutadores de red y el cableado dentro y entre los bastidores de Outposts también son totalmente redundantes. Un panel de conexiones de fibra proporciona conectividad entre un rack Outpost y la red local y sirve como punto de demarcación entre el entorno del centro de datos gestionado por el cliente y el entorno gestionado. AWS Outposts

Al igual que en la región, AWS es responsable de los servicios en la nube que se ofrecen en Outposts y asume responsabilidades adicionales a medida que selecciona e implementa servicios gestionados de nivel superior, como Amazon RDS en Outposts. Lea el Modelo de responsabilidad compartida de AWS y las páginas de preguntas frecuentes de cada servicio antes de elegir los servicios que quiera implementar en Outposts. Estos recursos proporcionan detalles adicionales sobre la división de responsabilidades entre usted y. AWS

Planteamiento en torno a los modos de falla

A la hora de diseñar una aplicación o un sistema de alta disponibilidad, se debe tener en cuenta qué componentes pueden fallar, qué impacto tendrán los errores de los componentes en el sistema y qué mecanismos se pueden implementar para mitigar o eliminar el impacto de estos errores. ¿La aplicación en cuestión se ejecuta en un único servidor, en un único rack o en un único centro de datos? ¿Qué ocurrirá cuando un servidor, rack o centro de datos experimente un error temporal o permanente? ¿Qué ocurre cuando se produce un error en un subsistema esencial como la red o en la propia aplicación? Hablamos de modos de error.

El usuario debe tener en cuenta los modos de error especificados en esta sección cuando planifique las implementaciones de Outposts y otras aplicaciones. En las secciones siguientes, se analizará cómo mitigar estos modos de error para proporcionar un mayor nivel de alta disponibilidad para el entorno de aplicaciones.

Modo de error 1: red

Una implementación de Outposts depende de una conexión resiliente con su región principal para su propia administración y supervisión. Las interrupciones de la red pueden deberse a diversos problemas, como errores del operador, errores del equipo e interrupciones del proveedor de servicios. Una implementación de Outposts, que puede estar compuesta por uno o más racks conectados entre sí en el sitio, se considera desconectada cuando no puede comunicarse con la región a través del enlace de servicio.

Las rutas de red redundantes pueden ayudar a mitigar el riesgo de que se produzcan eventos de desconexión. Se deben asignar el tráfico de red y las dependencias de la aplicación para conocer el impacto que los eventos de desconexión tendrían en las operaciones de las cargas de trabajo. El usuario debe planificar una redundancia de red suficiente para satisfacer los requisitos de disponibilidad de las aplicaciones.

Durante un evento de desconexión, las instancias que se ejecutan en una implementación de Outposts siguen ejecutándose y se puede acceder a ellas desde las redes en las instalaciones a través de la puerta de enlace local de Outposts (LGW). Las cargas de trabajo y los servicios locales pueden verse dañados o fallar si dependen de los servicios de la región. Las solicitudes de mutación (como iniciar o detener instancias en el Outpost), las operaciones del plano de control y la telemetría del servicio (por ejemplo, CloudWatch las métricas) fallarán mientras el Outpost esté desconectado de la región.

Modo de error 2: instancias

Las instancias de EC2 pueden verse afectadas o fallar si el servidor en el que se ejecutan tiene algún problema o si la instancia experimenta un error en el sistema operativo o en la aplicación. La forma en que las aplicaciones gestionan estos tipos de errores depende de la arquitectura de la aplicación. Las aplicaciones monolíticas suelen utilizar funciones de aplicaciones o sistemas para la recuperación, mientras que las arquitecturas modulares orientadas a los servicios o de microservicios suelen sustituir los componentes con errores para garantizar la disponibilidad del servicio.

Puede sustituir las instancias con errores por instancias nuevas mediante mecanismos automatizados, como los grupos de escalado automático de EC2. La recuperación automática de instancias puede reiniciar las instancias que fallan debido a errores en el servidor, siempre que haya suficiente capacidad libre disponible en los servidores restantes.

Modo de error 3: computación

Los servidores pueden fallar o verse dañados. Es posible que sea necesario ponerlos fuera de servicio (temporal o permanentemente) por diversos motivos, como errores de los componentes y operaciones de mantenimiento programadas. La forma en que los servicios del rack de Outposts gestionan los errores y los problemas de los servidores varía y puede depender de la forma en que los clientes configuren las opciones de alta disponibilidad.

El usuario debe solicitar una capacidad informática suficiente para admitir un modelo de disponibilidad `N+M`, en el que `N` es la capacidad requerida y `M` es la capacidad de reserva que se aprovisiona para adaptarse a los errores de los servidores.

Los reemplazos de hardware para los servidores averiados se proporcionan como parte de un servicio en rack totalmente gestionado. AWS Outposts AWS supervisa activamente el estado de todos los servidores y dispositivos de red en una implementación de Outpost. Si es necesario realizar un mantenimiento físico, AWS programará una visita al sitio del usuario para sustituir los componentes con errores. El aprovisionamiento de la capacidad de reserva permite mantener las cargas de trabajo en funcionamiento mientras los servidores con errores están fuera de servicio y en proceso de sustitución.

Modo de error 4: racks o centros de datos

Los errores de los racks pueden producirse debido a una pérdida total de la alimentación de estos o a problemas con el entorno, como una pérdida de la refrigeración o daños físicos en el centro

de datos a causa de una inundación o un terremoto. Las deficiencias en las arquitecturas de distribución eléctrica del centro de datos o los errores que se producen durante el mantenimiento de la alimentación estándar del centro de datos pueden provocar la pérdida de energía en uno o más racks, o incluso en todo el centro de datos.

Estos escenarios se pueden mitigar mediante la implementación de la infraestructura en varios pisos o ubicaciones del centro de datos que sean independientes entre sí dentro del mismo campus o área metropolitana.

Adoptar este enfoque con AWS Outposts rack requerirá una cuidadosa consideración de cómo se diseñan y distribuyen las aplicaciones para que se ejecuten en varios Outposts lógicos independientes a fin de mantener la disponibilidad de las aplicaciones.

Modo de error 5: zona o AWS región de disponibilidad

Cada implementación de Outposts está anclada a una zona de disponibilidad (AZ) específica dentro de una Región de AWS. Los errores de la región principal o la zona de disponibilidad de anclaje pueden provocar la pérdida de la gestión y la mutabilidad de Outposts, así como interrumpir la comunicación de red entre Outposts y la región.

Al igual que los errores de la red, los de la zona de disponibilidad o la región pueden provocar que Outposts se desconecte de la región. Las instancias que se ejecutan en una implementación de Outposts siguen ejecutándose y se puede acceder a ellas desde las redes en las instalaciones a través de la puerta de enlace local (LGW) de Outposts. Además, si dependen de los servicios de la región, pueden verse dañadas o fallar, tal y como se ha descrito anteriormente.

Para mitigar el impacto de los fallos en las AWS zonas de disponibilidad y las regiones, puedes desplegar varios Outposts, cada uno de ellos anclado a una zona de disponibilidad o región diferente. A continuación, puede diseñar su carga de trabajo para que funcione en un modelo de despliegue distribuido de varios Outpost utilizando muchos de los mecanismos y patrones arquitectónicos similares a los que utiliza actualmente para diseñar e implementar. AWS

Creación de aplicaciones de alta disponibilidad y soluciones de infraestructura con AWS Outposts rack

Con AWS Outposts rack, puede crear, administrar y escalar aplicaciones locales de alta disponibilidad utilizando herramientas y servicios en AWS la nube que ya conoce. Es importante entender que, por lo general, las arquitecturas y los enfoques de alta disponibilidad en la nube difieren de las arquitecturas de alta disponibilidad tradicionales en las instalaciones que podrían estar actualmente en uso en su centro de datos.

Con las implementaciones tradicionales en las instalaciones de aplicaciones de alta disponibilidad, las aplicaciones se implementan en máquinas virtuales (VM). Para garantizar el buen funcionamiento y estado de estas máquinas virtuales, se implementan y mantienen sistemas e infraestructuras de TI complejos. Las VM suelen tener identidades específicas y cada una de ellas puede desempeñar un rol fundamental en la arquitectura total de la aplicación.

Estos roles de la arquitectura se encuentran estrechamente acoplados a las identidades de las VM. Los arquitectos de sistemas aprovechan las características de la infraestructura de TI para brindar entornos de tiempo de ejecución de VM de alta disponibilidad que, a su vez, proporcionan a cada una de ellas un acceso fiable a la capacidad informática, los volúmenes de almacenamiento y los servicios de red. Si una VM falla, se ejecutan procesos de recuperación automatizados o manuales para restaurar la VM con errores a un buen estado, a menudo en otra infraestructura o en un centro de datos completamente diferente.

Las arquitecturas de alta disponibilidad en la nube adoptan un enfoque diferente. AWS los servicios en la nube proporcionan capacidades confiables de cómputo, almacenamiento y redes. Los componentes de la aplicación se implementan en instancias de EC2, contenedores, funciones sin servidor u otros serviios administrados.

Una instancia es una instanciación de un componente de una aplicación (quizás uno de los muchos que desempeñan ese rol). Los componentes de la aplicación se acoplan de forma flexible entre sí y al rol que desempeñan en la arquitectura total de la aplicación. La identidad individual de una instancia no suele tener importancia. Se pueden crear o destruir instancias adicionales para escalarlas o reducirlas verticalmente según la demanda. Las instancias fallidas o en mal estado se sustituyen por instancias nuevas en buen estado.

AWS Outposts rack es un servicio totalmente gestionado que extiende la AWS computación, el almacenamiento, las redes, las bases de datos y otros servicios en la nube a las ubicaciones

locales para ofrecer una experiencia híbrida verdaderamente coherente. No considere el servicio de bastidores de Outposts un sustituto directo de los sistemas de infraestructura de TI con mecanismos de alta disponibilidad tradicionales en las instalaciones. Intentar utilizar AWS los servicios y Outposts para dar soporte a una arquitectura de alta disponibilidad local tradicional va en contra de los patrones.

Las cargas de trabajo que se ejecutan en AWS Outposts rack utilizan mecanismos de alta disponibilidad en la nube, como Amazon EC2 Auto Scaling (para escalar horizontalmente y satisfacer las demandas de carga de trabajo), las comprobaciones de estado de EC2 (para detectar y eliminar las instancias en mal estado) y los balanceadores de carga de aplicaciones (para redirigir el tráfico de carga de trabajo entrante a instancias escaladas o reemplazadas). Al migrar aplicaciones a la nube, ya sea a un AWS Outposts rack Región de AWS o a un rack, debe actualizar la arquitectura de las aplicaciones de alta disponibilidad para empezar a aprovechar los servicios gestionados en la nube y los mecanismos de alta disponibilidad en la nube.

En las siguientes secciones, se presentan los patrones de arquitectura, los antipatrones y las prácticas recomendadas para implementar el AWS Outposts rack en sus entornos locales a fin de ejecutar cargas de trabajo con requisitos de alta disponibilidad. Estas secciones ofrecen una introducción sobre los patrones y prácticas; sin embargo, no proporcionan detalles en torno a la configuración y la implementación. El usuario debe leer y familiarizarse con las preguntas frecuentes y la guía del usuario de los bastidores de AWS Outposts , así como las preguntas frecuentes y la documentación de los servicios que se ejecutan en los bastidores de Outposts, a medida que prepara el entorno para el bastidor de Outposts y las aplicaciones para la migración a los servicios de AWS .

Temas

Red

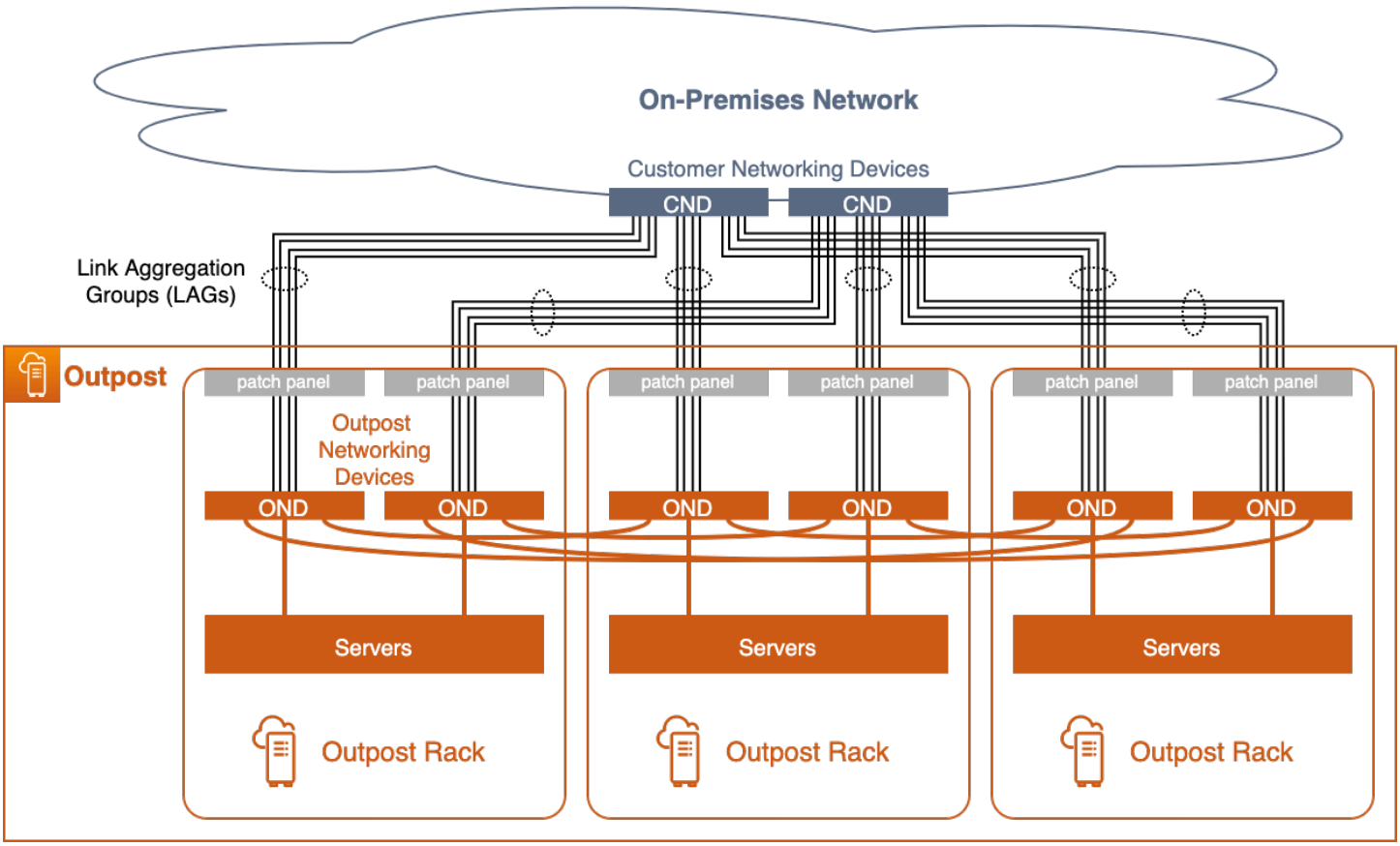
Para que las operaciones de administración, supervisión y servicio funcionen correctamente, la implementación de una instancia de Outposts depende de una conexión resiliente a su zona de disponibilidad (AZ) de anclaje. Debe aprovisionar su red local para proporcionar conexiones de red

redundantes para cada rack de Outpost y una conectividad fiable hasta los puntos de anclaje de la nube. AWS También deben tenerse en cuenta las rutas de red entre las cargas de trabajo de las aplicaciones que se ejecutan en Outposts y el resto de sistemas en las instalaciones y la nube con los que se comunican. ¿Cómo se va a enrutar este tráfico en la red?

Temas

Conexión de redes

Cada AWS Outposts rack está configurado con top-of-rack conmutadores redundantes denominados dispositivos de red Outpost (OND). Los servidores informáticos y de almacenamiento de cada bastidor se conectan a ambos OND. Cada OND se debe conectar a un conmutador independiente denominado dispositivo de red del cliente (CND) del centro de datos del usuario para proporcionar diversas rutas físicas y lógicas para cada bastidor de Outposts. Los OND se conectan a los CND mediante una o más conexiones físicas, con cables de fibra óptica y transceptores ópticos. Las conexiones físicas se configuran en enlaces lógicos de grupos de agregación de enlaces (LAG).



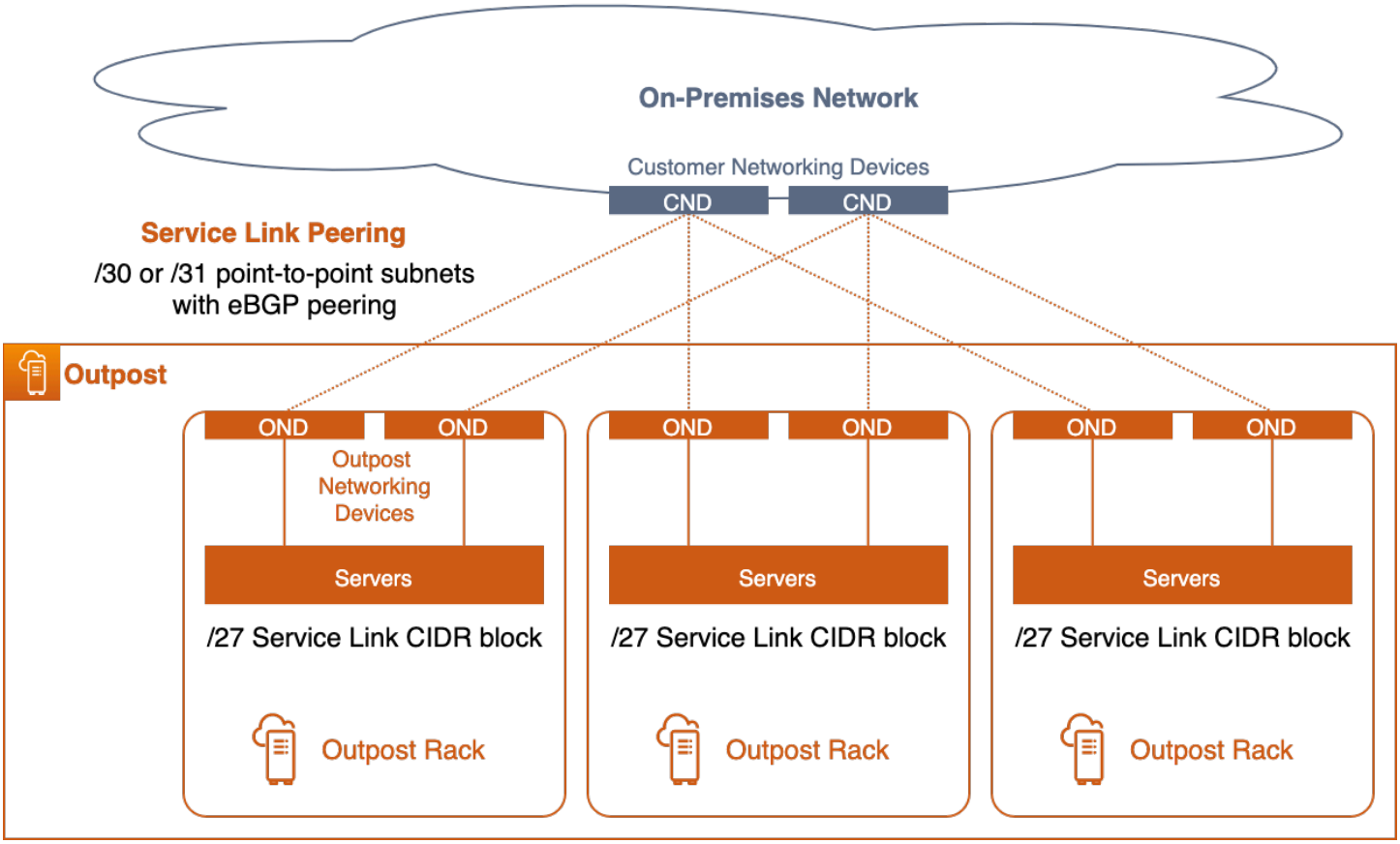
Instancia de múltiples bastidores de Outposts con conexiones redundantes de red

Los enlaces OND a CND siempre se configuran en un LAG, aunque la conexión física sea un solo cable de fibra óptica. La configuración de los enlaces como grupos LAG permite aumentar el ancho de banda de los enlaces mediante la incorporación de conexiones físicas adicionales al grupo lógico. Los enlaces LAG se configuran como redes troncales Ethernet de estándar IEEE 802.1q para permitir la creación de redes segregadas entre Outposts y la red en las instalaciones.

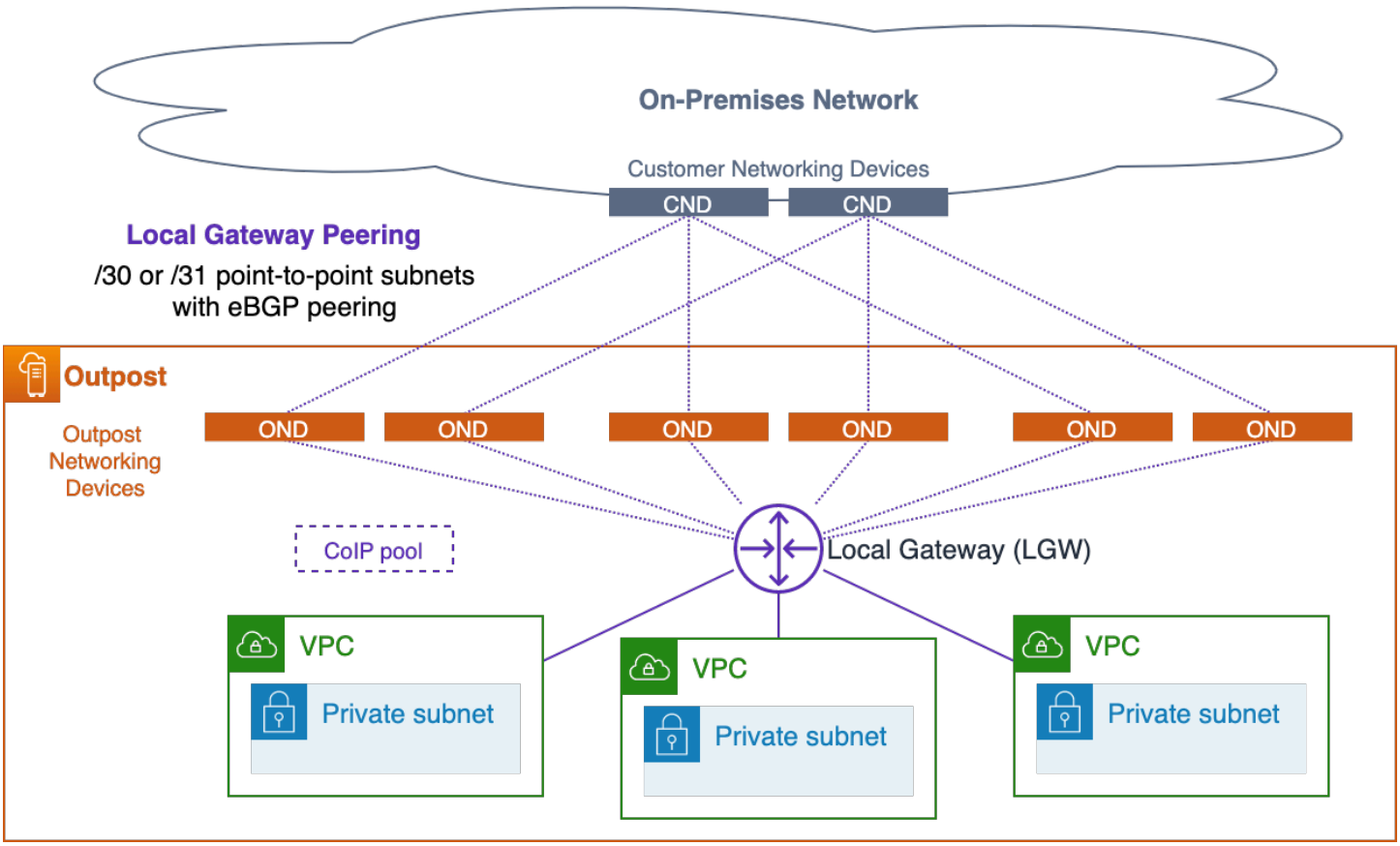
Cada instancia de Outposts tiene al menos dos redes segregadas de forma lógica que deben comunicarse con la red del cliente o a través de ella:

- Red de enlace de servicio: asigna las direcciones IP del enlace de servicio a los servidores de Outpost y facilita la comunicación con la red local para permitir que los servidores se conecten de nuevo a los puntos de anclaje de Outpost en la región.
- Red de puerta de enlace local: permite la comunicación entre las subredes de VPC de Outposts y la red en las instalaciones a través de la puerta de enlace local (LGW) de Outposts.

Estas redes segregadas se conectan a la red local mediante un conjunto de conexiones IP a través de los enlaces LAG. point-to-point Cada enlace LAG de OND a CND se configura con identificadores de VLAN, subredes IP point-to-point (/30 o /31) e interconexión eBGP para cada red segregada (enlace de servicio y LGW). Debe considerar los enlaces LAG, con sus point-to-point VLAN y subredes, como conexiones de capa 3 enrutadas y segmentadas de capa 2. Las conexiones IP enrutadas proporcionan rutas lógicas redundantes que facilitan la comunicación entre las redes segregadas de Outposts y la red en las instalaciones.



Emparejamiento de enlaces de servicio



Interconexión de una puerta de enlace local

Los enlaces LAG de capa 2 (y sus VLAN) deben terminar en los conmutadores CND conectados directamente; por su parte, las interfaces IP y la interconexión BGP deben configurarse en los conmutadores CND. Las VLAN de los enlaces LAG no deben servir de puente entre los distintos conmutadores del centro de datos. Para obtener más información, consulte Conectividad de la capa de red en la Guía del usuario de AWS Outposts .

Dentro de un Outpost lógico con varios racks, los OND están interconectados de forma redundante para proporcionar una conectividad de red de alta disponibilidad entre los racks y las cargas de trabajo que se ejecutan en los servidores. AWS es responsable de la disponibilidad de la red en el Outpost.

Prácticas recomendadas para conexiones de redes de alta disponibilidad

- Cada dispositivo de red de Outposts (OND) de un bastidor de Outposts debe conectarse a un dispositivo de red del cliente (CND) independiente del centro de datos.

- Los enlaces de capa 2, las VLAN, las subredes IP de capa 3 y la interconexión BGP deben terminar en los conmutadores de los dispositivos de red del cliente (CND) conectados directamente. Las VLAN de OND a CND no deben conectarse entre varios CND o a través de la red en las instalaciones.
- Deben añadirse enlaces a los grupos de agregación de enlaces (LAG) para aumentar el ancho de banda disponible entre Outposts y el centro de datos. No debe confiarse en el ancho de banda agregado de las diversas rutas que atraviesan ambos OND.
- Pueden utilizarse las diversas rutas que atraviesan los OND redundantes para proporcionar una conectividad resiliente entre las redes de Outposts y la red en las instalaciones.
- Para lograr una redundancia óptima y permitir un mantenimiento de los OND sin interrupciones, recomendamos que los clientes configuren los anuncios y las políticas de BGP de la siguiente manera:
  - El equipo de red del cliente debe recibir anuncios de BGP de Outposts sin cambiar los atributos de BGP y debe habilitar la opción de varias rutas y el balanceo de carga de BGP para lograr flujos de tráfico entrante óptimos (del cliente a Outposts). Los prefijos BGP de Outposts utilizan el atributo AS-Path para desviar el tráfico de un OND o un enlace ascendente concreto en caso de que sea necesario realizar tareas de mantenimiento. La red del cliente debería preferir las rutas de Outposts con una longitud del atributo AS-Path de 1 a las rutas con una longitud del atributo AS-Path de 4; es decir, debe reaccionar al atributo AS-Path.
  - La red del cliente debe anunciar prefijos BGP iguales, con los mismos atributos en todos los OND de Outposts. De forma predeterminada, la carga de red de Outposts equilibra el tráfico saliente (hacia el cliente) de todos los enlaces ascendentes. Las políticas de enrutamiento se utilizan en Outposts para desviar el tráfico de un OND en particular en caso de que sea necesario realizar tareas de mantenimiento. Para modificar el tráfico y realizar el mantenimiento de forma no disruptiva, se requieren prefijos BGP iguales por parte del cliente en todos los OND. Cuando sea necesario realizar tareas de mantenimiento en la red del cliente, recomendamos utilizar prefijos con el atributo AS-Path a fin de desviar temporalmente el tráfico procedente de un enlace ascendente o dispositivo concreto.

Conectividad de anclaje

Un enlace de servicio de Outpost se conecta a puntos de anclaje públicos o privados (no a ambos) en una zona de disponibilidad (AZ) específica de la región principal del Outpost. Los servidores de Outpost inician las conexiones VPN de enlace de servicio saliente desde sus direcciones IP de enlace de servicio hasta los puntos de anclaje de la AZ de anclaje. Estas conexiones utilizan los

puertos UDP y TCP 443. AWS es responsable de la disponibilidad de los puntos de anclaje en la Región.

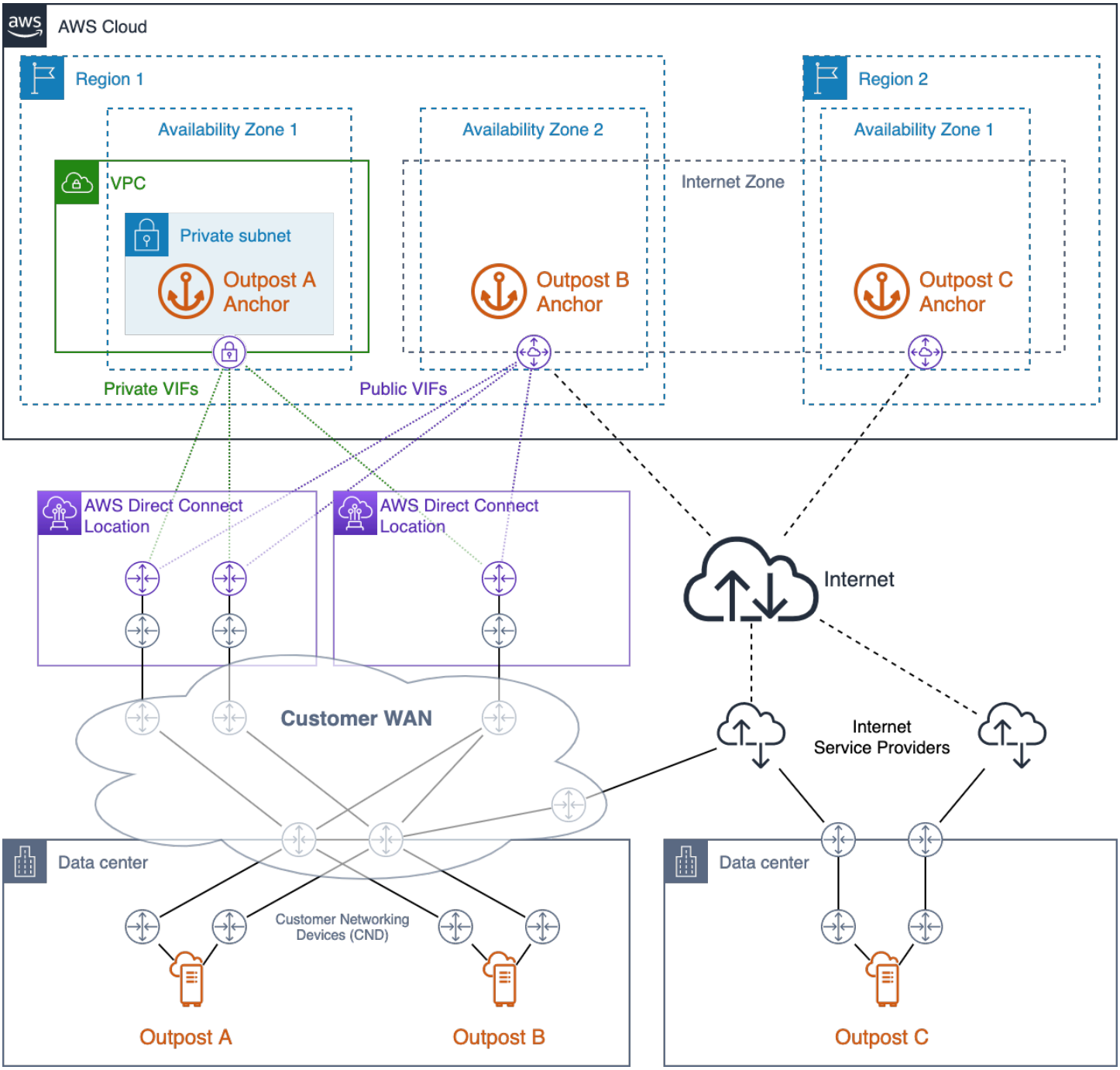
Debe asegurarse de que las direcciones IP del enlace del servicio Outpost puedan conectarse a través de su red a los puntos de anclaje de la zona de anclaje. Las direcciones IP del enlace de servicio no necesitan comunicarse con otros hosts de la red local.

Los puntos de anclaje públicos residen en los rangos de direcciones IP públicas de la región (en los bloques CIDR del servicio EC2) y se puede acceder a ellos a través de Internet o de interfaces virtuales públicas (VIF) de AWS Direct Connect (DX). El uso de puntos de anclaje públicos permite una selección de rutas más flexible, ya que el tráfico del enlace de servicio se puede enrutar a través de cualquier ruta disponible que pueda llegar correctamente a los puntos de anclaje de la Internet pública.

Los puntos de anclaje privados permiten usar al usuario sus propios rangos de direcciones IP para establecer la conectividad de anclaje. Los puntos de anclaje privados se crean en una subred privada dentro de una VPC dedicada mediante direcciones IP asignadas por el cliente. La VPC se crea en la VPC propietaria del recurso Outpost y usted es responsable de garantizar Cuenta de AWS que la VPC esté disponible y configurada correctamente (¡no la elimine!). Se debe acceder a los puntos de anclaje privados mediante VIF privadas de Direct Connect.

Se deben aprovisionar rutas de red redundantes entre Outposts y los puntos de anclaje de la región, con conexiones que terminen en dispositivos independientes de más de una ubicación. Se debe configurar el direccionamiento dinámico para redirigir automáticamente el tráfico a rutas alternativas cuando las conexiones o los dispositivos de red fallen. Se debe aprovisionar una capacidad de red suficiente para garantizar que un error en una ruta WAN no sobrecargue las rutas restantes.

El siguiente diagrama muestra tres Outposts con rutas de red redundantes a sus AZ de anclaje, AWS Direct Connect además de conectividad pública a Internet. Las instancias de Outposts A y B están ancladas a diferentes zonas de disponibilidad de la misma región. La instancia de Outposts A se conecta a puntos de anclaje privados en la AZ 1 de la región 1. La instancia de Outposts B se conecta a puntos de anclaje públicos en la AZ 2 de la región 1. La instancia de Outposts B se conecta a puntos de anclaje públicos en la AZ 2 de la región 1.



Conectividad de anclaje de alta disponibilidad con AWS Direct Connect acceso público a Internet

La instancia de Outposts A tiene tres rutas de red redundantes para llegar al punto de anclaje privado. Hay dos rutas disponibles a través de circuitos de Direct Connect redundantes en una única ubicación de Direct Connect. La tercera ruta está disponible a través de un circuito de Direct Connect en una segunda ubicación de Direct Connect. Este diseño mantiene el tráfico del enlace de servicio

del Outpost A en las redes privadas y proporciona una redundancia de rutas que permite el fallo de cualquiera de los circuitos de Direct Connect o el fallo de toda una ubicación de Direct Connect.

La instancia de Outposts B tiene cuatro rutas de red redundantes para llegar al punto de anclaje privado. Tres rutas están disponibles a través de VIF públicas aprovisionadas en los circuitos y ubicaciones de Direct Connect que utiliza la instancia de Outposts A. La cuarta ruta está disponible a través de la WAN del cliente y la Internet pública. El tráfico del enlace de servicio de Outpost B se puede enrutar a través de cualquier ruta disponible que pueda llegar correctamente a los puntos de anclaje de la Internet pública. El uso de las rutas Direct Connect puede proporcionar una latencia más coherente y una mayor disponibilidad de ancho de banda, mientras que la ruta de Internet pública se puede usar para la recuperación de desastres o aumentar el ancho de banda.

La instancia de Outposts C tiene dos rutas de red redundantes para llegar al punto de anclaje privado. La instancia de Outposts C se encuentra implementada en un centro de datos diferente al de las instancias de Outposts A y B. El centro de datos de la instancia de Outposts C no tiene circuitos dedicados que se conecten a la WAN del cliente. En cambio, el centro de datos tiene conexiones a Internet redundantes proporcionadas por dos proveedores de servicios de Internet (ISP) diferentes. El tráfico del enlace de servicio de Outpost C puede enrutarse a través de cualquiera de las redes del ISP para llegar a los puntos de anclaje de la Internet pública. Este diseño ofrece flexibilidad para enrutar el tráfico de los enlaces de servicio a través de cualquier conexión pública a Internet disponible. Sin embargo, la end-to-end ruta depende de las redes públicas de terceros, donde la disponibilidad del ancho de banda y la latencia de la red fluctúan.

La ruta de red entre un Outpost y sus puntos de anclaje de enlace de servicio debe cumplir con la siguiente especificación de ancho de banda:

- 500 Mbps: 1 Gbps de ancho de banda disponible por bastidor de Outposts (por ejemplo, para 3 bastidores, el ancho de banda disponible debe ser de entre 1,5 y 3 Gbps)

Prácticas recomendadas para una conectividad de anclaje de alta disponibilidad

- Proporcione rutas de red redundantes entre cada implementación de Outposts y sus puntos de anclaje en la región.
- Utilice las rutas de Direct Connect (DX) para controlar la latencia y la disponibilidad del ancho de banda.
- Asegúrese de que los puertos TCP y UDP 443 estén abiertos (salida) desde los bloques CIDR de los enlaces de servicio de Outposts hasta los rangos de direcciones IP de EC2 de la región principal. Confirme que los puertos estén abiertos en todas las rutas de red.

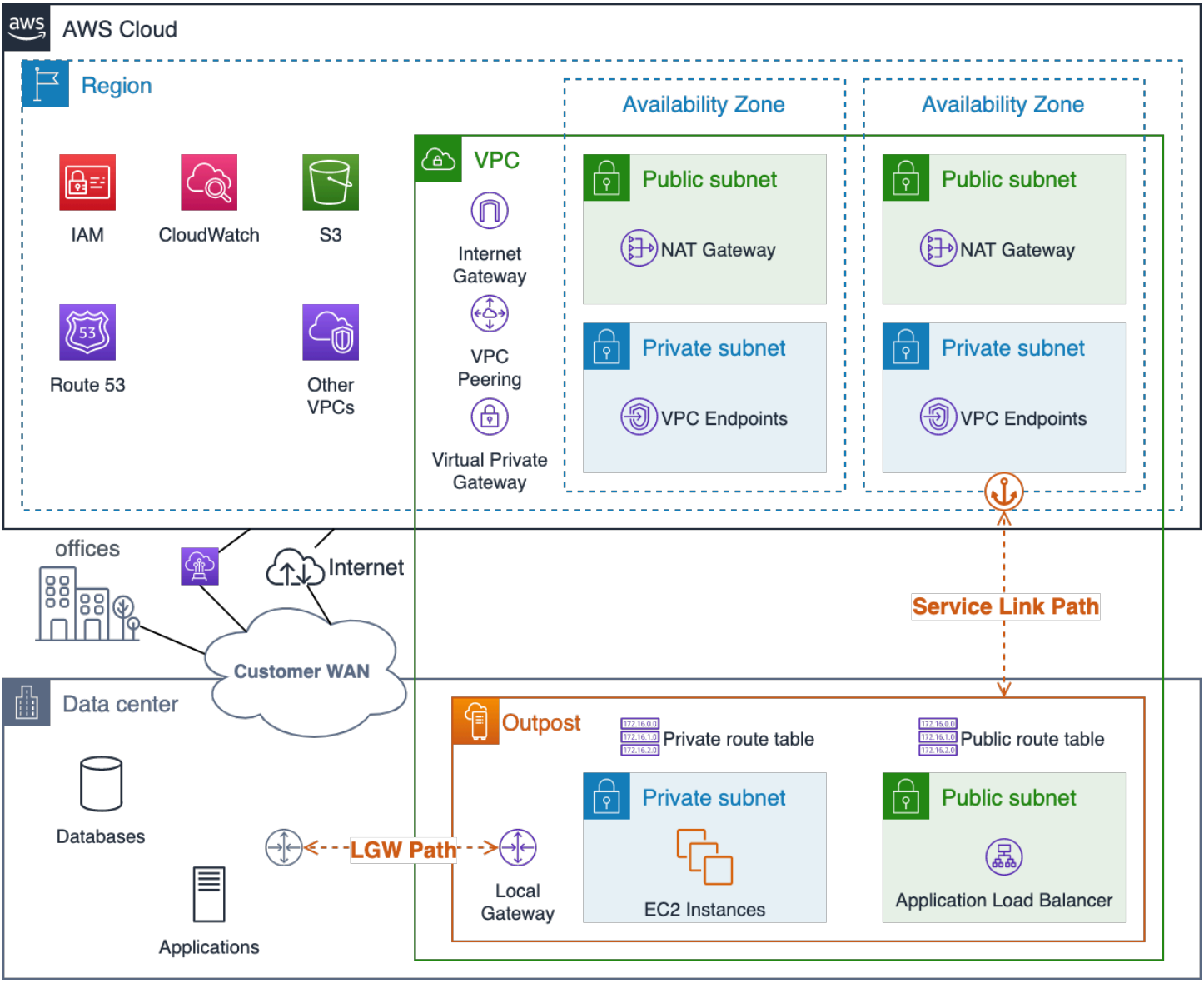
- Compruebe que cada ruta cumple con los requisitos de latencia y disponibilidad de ancho de banda específicos.
- Utilice el direccionamiento dinámico para automatizar el redireccionamiento del tráfico en caso de producirse errores en la red.
- Pruebe el enrutamiento del tráfico del enlace de servicio a través de cada ruta de red planificada para asegurarse de que la ruta funcione según lo esperado.

Enrutamiento de aplicaciones y cargas de trabajo

Hay dos rutas de salida de Outposts para las cargas de trabajo de las aplicaciones:

- La ruta del enlace de servicio
- La ruta de la puerta de enlace local (LGW)

Las tablas de enrutamiento de la subred de Outposts se configuran para controlar la ruta que se debe seguir para llegar a las redes de destino. Las rutas que apuntan a la LGW dirigirán el tráfico desde la puerta de enlace local hacia la red en las instalaciones. Las rutas que apuntan a los servicios y recursos de la región, como Internet Gateway, NAT Gateway, Virtual Private Gateway y TGW, utilizarán Service Link para alcanzar estos objetivos. Si tienes una conexión de emparejamiento de VPC con varias VPC en el mismo Outpost, el tráfico entre las VPC permanece en el Outpost y no utiliza el enlace de servicio para volver a la región. Para obtener información sobre la interconexión de VPC, consulte Conectar VPC mediante la interconexión de VPC en la Guía del usuario de Amazon VPC.



Visualización del enlace del servicio Outpost y de las rutas de red LGW

A la hora de planificar el enrutamiento de las aplicaciones, se debe tener en cuenta tanto el funcionamiento normal como la disponibilidad limitada del servicio y el enrutamiento cuando se producen errores en la red. La ruta de enlace de servicio no está disponible si Outposts está desconectado de la región.

Se deben aprovisionar diversas rutas y configurar el direccionamiento dinámico entre la LGW de Outposts y las aplicaciones, sistemas y usuarios esenciales en las instalaciones. Las rutas de red redundantes permiten a la red redirigir el tráfico en caso de error y garantizar que los recursos en las instalaciones puedan comunicarse con las cargas de trabajo que se ejecutan en Outposts en caso de que la red falle parcialmente.

Las configuraciones de enrutamiento de las VPC de Outposts son estáticas. Las tablas de enrutamiento de subred se configuran mediante la CLI AWS Management Console, las API y otras herramientas de infraestructura como código (IaC); sin embargo, no podrá modificar las tablas de enrutamiento de subred durante un evento de desconexión. Deberá restablecerse la conectividad entre Outposts y la región para que las tablas de enrutamiento puedan actualizarse. Deben usarse las mismas rutas para las operaciones normales que las previstas para los eventos de desconexión.

Los recursos del Outpost pueden acceder a Internet a través del enlace de servicio y una puerta de enlace de Internet (IGW) en la región o a través de la ruta de puerta de enlace local (LGW). Enrutar el tráfico de Internet a través de la ruta LGW y la red local permite utilizar los puntos de entrada y salida de Internet locales existentes y puede ofrecer una latencia más baja, MTU más altas y costes de salida de AWS datos más bajos en comparación con el uso de la ruta de enlace de servicio a una IGW de la región.

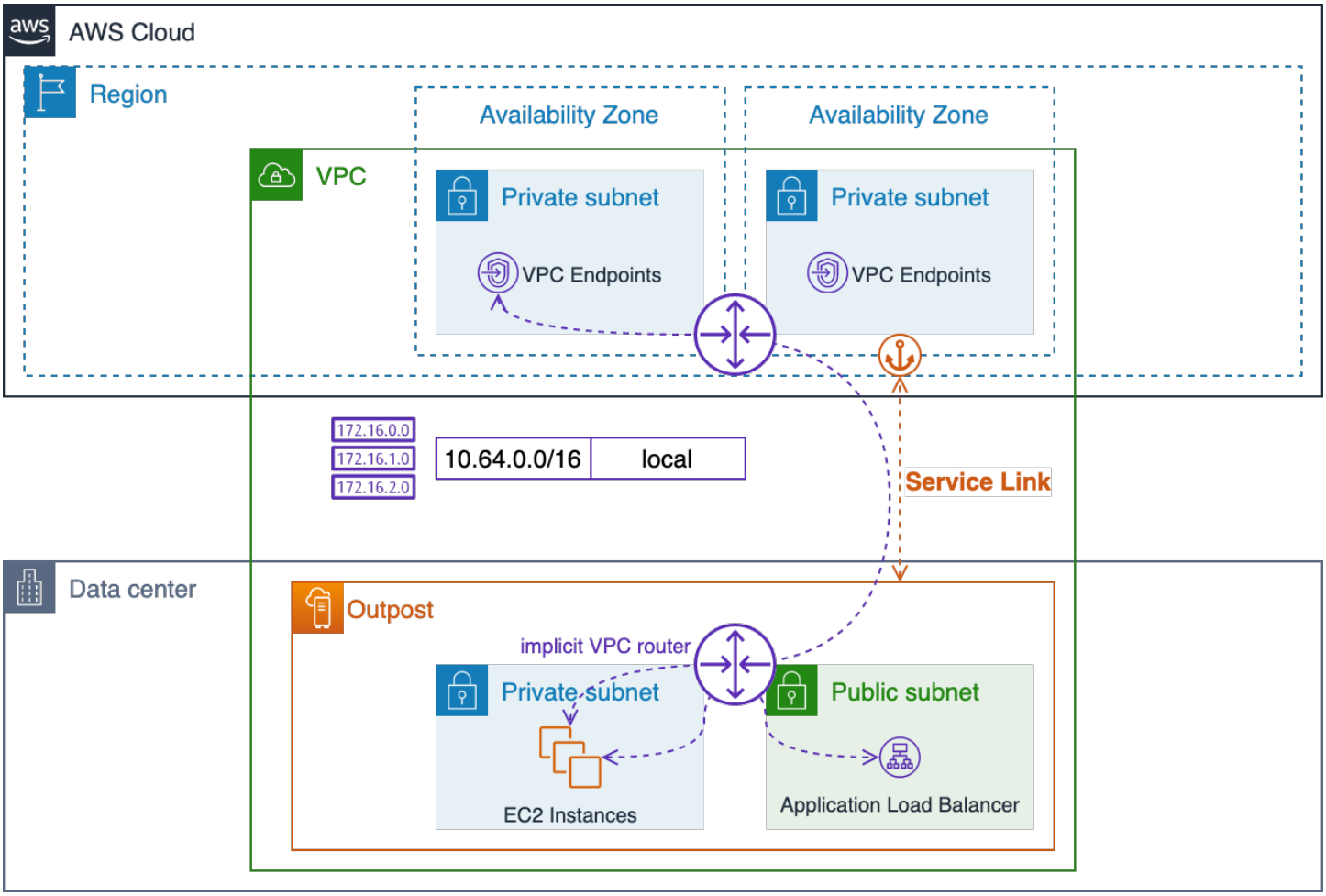
Si la aplicación debe ejecutarse en las instalaciones y es necesario que se pueda acceder a ella desde la Internet pública, el tráfico de la aplicación se debe enrutar a través de las conexiones a Internet en las instalaciones hasta la LGW con el objetivo de llegar a los recursos de Outposts.

Si bien se pueden configurar las subredes en Outposts como subredes públicas de la región, no representa la práctica más deseable para la mayoría de los casos de uso. El tráfico de Internet entrante entrará a través del enlace de servicio Región de AWS y se enrutará a través del enlace de servicio hasta los recursos que se encuentran en el Outpost.

El tráfico de respuesta, a su vez, se enrutará a través del enlace del servicio y regresará a través de las conexiones a Internet del Región de AWS servicio. Este patrón de tráfico puede aumentar la latencia e incurrir en gastos de salida de datos cuando el tráfico salga de la región en dirección a Outposts y el tráfico de retorno vuelva a través de la región hacia Internet. Si la aplicación se puede ejecutar en la región, será el mejor lugar para ejecutarla.

El tráfico entre los recursos de la VPC (en la misma VPC) siempre seguirá la ruta CIDR de la VPC local y los enrutadores de VPC implícitos lo redirigirán entre las subredes.

Por ejemplo, el tráfico entre una instancia EC2 que se ejecuta en Outpost y un punto final de VPC en la región siempre se enrutará a través del enlace de servicio.



Enrutamiento de la VPC local a través de enrutadores implícitos

Prácticas recomendadas para el enrutamiento de aplicaciones y cargas de trabajo

- Siempre que sea posible, utilice la ruta de la puerta de enlace local (LGW) en lugar de la ruta del enlace de servicio.
- Enrute el tráfico de Internet a través de la ruta LGW.
- Configure las tablas de enrutamiento de la subred de Outposts con un conjunto estándar de rutas; se usarán tanto para las operaciones normales como durante los eventos de desconexión.
- Aprovisione rutas de red redundantes entre la LGW de Outposts y los recursos esenciales de las aplicaciones en las instalaciones. Utilice el direccionamiento dinámico para automatizar el redireccionamiento del tráfico en caso de producirse errores en la red en las instalaciones.

Cálculo

Si bien la capacidad de entrada de Amazon EC2 Regiones de AWS es aparentemente infinita, la capacidad de Outposts es finita. El usuario es responsable de planificar y administrar la capacidad informática de las implementaciones de Outposts.

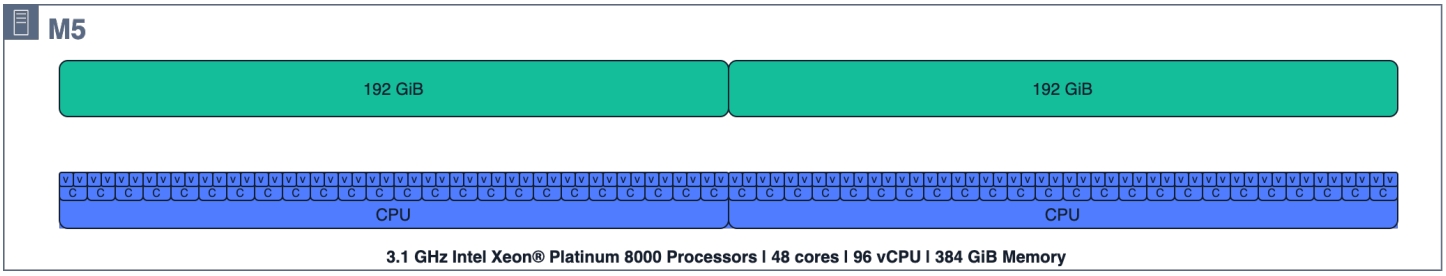
Temas

Planificación de la capacidad

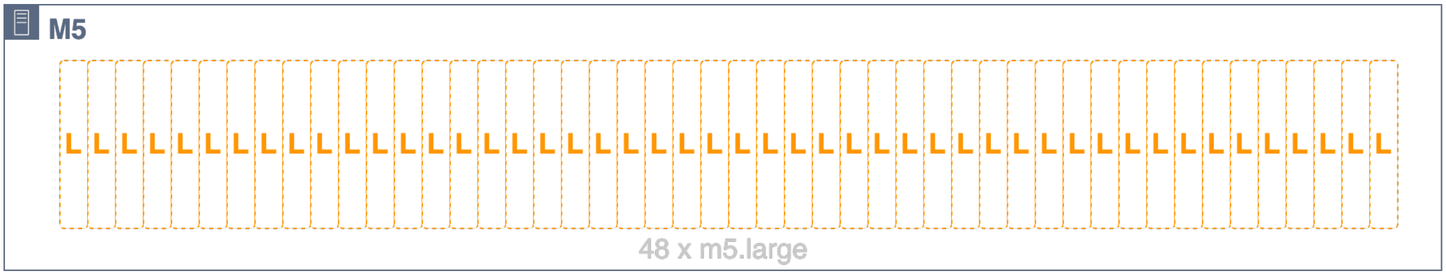
Si bien la capacidad de Amazon EC2 Regiones de AWS es aparentemente infinita, la capacidad de Outposts es finita, limitada por el volumen total de capacidad de cómputo solicitada. El usuario es responsable de planificar y administrar la capacidad informática de las implementaciones de Outposts. El usuario debe solicitar una capacidad informática suficiente para admitir un modelo de disponibilidad N+M, en el que N es la capacidad requerida y M es el número de servidores de reserva aprovisionados para adaptarse a los errores de los servidores. N+1 y N+2 son los niveles de disponibilidad más comunes.

Cada servidor (`C5`,, `M5R5`, etc.) admite una sola familia de instancias EC2. Antes de lanzar instancias en servidores informáticos de EC2, debe proporcionar diseños de ranuras que especifiquen los tamaños de instancia de EC2 que desea que proporcione cada servidor. AWS configura cada servidor con el diseño de ranuras solicitado.

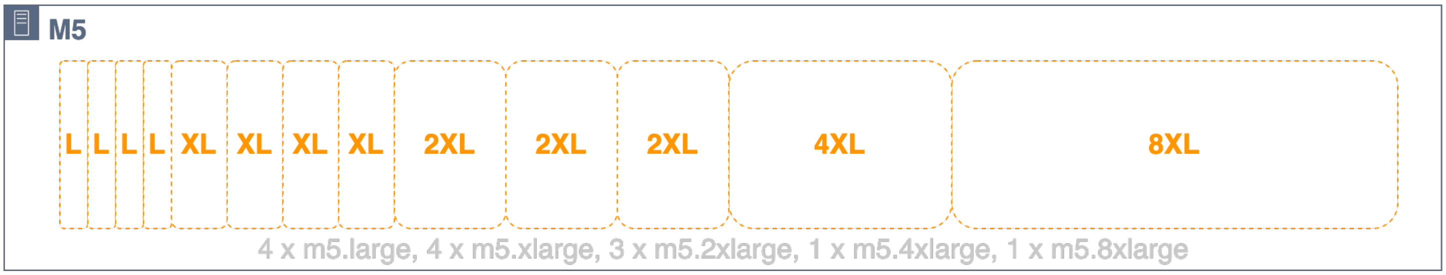
Los servidores pueden tener ranuras homogéneas cuando todas las ranuras tienen el mismo tamaño de instancia (por ejemplo, 48 `m5.large` ranuras) o heterogéneamente con una mezcla de tipos de instancias (por ejemplo, 4, 4`m5.large`, 3 `m5.xlarge` `m5.2xlarge`m5.4xlarge, 1 y 1`m5.8xlarge`). Consulte las tres figuras siguientes para ver una visualización de estas configuraciones de asignación de ranuras.



*m5.24xlarge*recursos informáticos del servidor



*m5.24xlarge*servidor distribuido homogéneamente en 48 ranuras *m5.large*



*m5.24xlarge*servidor distribuido de forma heterogénea en 4*m5.large*, 4, 3 *m5.xlargem5.2xlarge*, 1 y 1 ranuras *m5.4xlarge m5.8xlarge*

No es necesario asignar toda la capacidad del servidor a los slots. Se pueden añadir más slots a un servidor que tenga capacidad disponible sin asignar. Si se quiere modificar el diseño de la configuración de los slots, debe abrirse un ticket de soporte. Enterprise Support podría solicitar el cierre o reinicio de determinadas instancias para completar una solicitud de reasignación de slots si el nuevo diseño de los slots no se puede aplicar mientras determinados slots estén ocupados por instancias en ejecución.

Todos los servidores aportan slots aprovisionados a los grupos de capacidad de EC2 de Outposts, y todos los slots de un tipo y tamaño de instancia determinados se administran como un único grupo de capacidad de EC2. Por ejemplo, el servidor anterior con ranuras heterogéneas`m5.large`,, `m5.xlarge` `m5.2xlargem5.4xlarge`, y las ranuras proporcionaría estas `m5.8xlarge` ranuras a cinco grupos de capacidad de EC2, un grupo para cada tipo y tamaño de instancia.

Es importante tener en cuenta la distribución de los servidores y los grupos de capacidad de EC2 al planificar la capacidad sobrante para la disponibilidad de los servidores N+M. AWS detecta cuando un servidor falla o se degrada y programa una visita al sitio para reemplazar el servidor averiado. Los grupos de capacidades de EC2 deben diseñarse de manera que toleren que al menos un servidor de cada familia de instancias (N+1) de Outposts falle. Con este nivel mínimo de disponibilidad de los

servidores, cuando un servidor falla o es necesario dejarlo fuera de servicio, pueden reiniciarse las instancias con errores o un rendimiento reducido en los slots de reserva del resto de servidores de la misma familia.

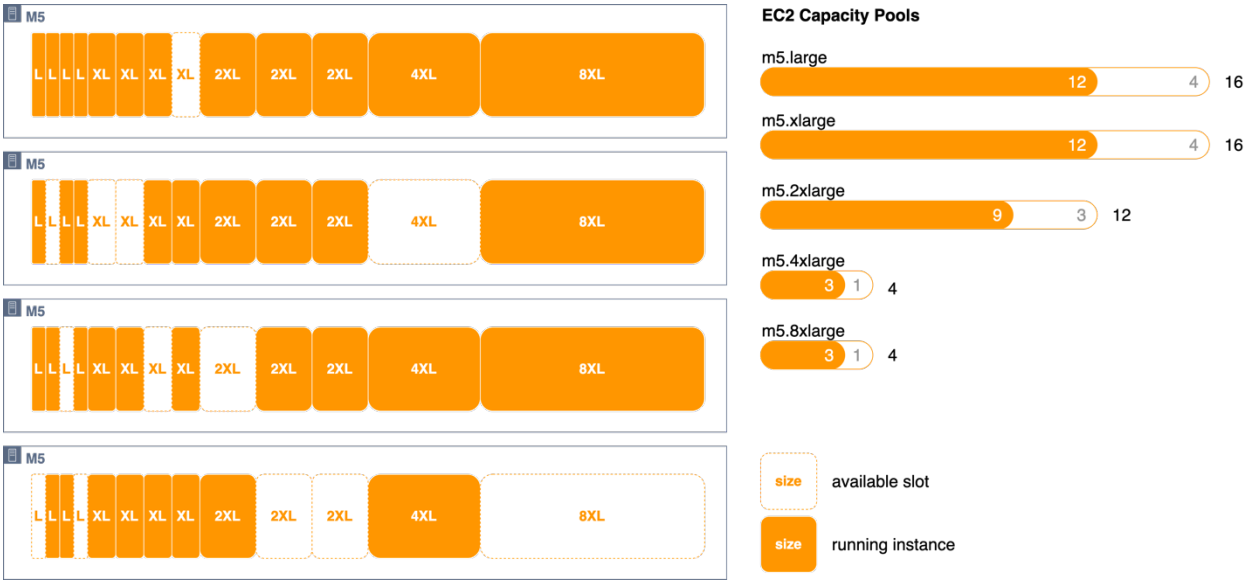
Planificar la disponibilidad de N+M es sencillo cuando se dispone de servidores con una configuración de slots homogénea o grupos de servidores con una configuración de slots homogénea y diseños idénticos. Solo se tiene que calcular la cantidad de servidores (N) que se necesita para ejecutar todas las cargas de trabajo y, a continuación, añadir (M) servidores adicionales para satisfacer los requisitos de disponibilidad de los servidores en caso de error o tareas de mantenimiento.

Las siguientes configuraciones de asignación de ranuras no se pueden utilizar debido a los límites de la NUMA:

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

Consulte a su Cuenta de AWS equipo para validar la configuración de ranuras de AWS Outposts estanterías planificada.

En la siguiente figura, cuatro `m5.24xlarge` servidores tienen ranuras heterogéneas con un diseño de ranuras idéntico. Los cuatro servidores crean cinco grupos de capacidad de EC2. Cada grupo se ejecuta con un uso máximo (75 %) para mantener una disponibilidad de N+1 para las instancias que se ejecutan en estos cuatro servidores. Si un servidor falla, hay espacio suficiente para reiniciar las instancias con errores en los servidores restantes.



Slots de los servidores de EC2, instancias en ejecución y grupos de slots

Para diseños de slots más complejos, en los que los servidores no tienen configuraciones idénticas, se tendrá que calcular la disponibilidad de N+M para cada grupo de capacidad de EC2. Se puede utilizar la siguiente fórmula para calcular cuántos servidores (que aportan slots a un grupo de capacidad de EC2 determinado) pueden fallar y, aun así, permitir que los servidores restantes alojen las instancias en ejecución:

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

Donde:

- $poolSlots_{available}$ es el número de slots disponibles en un grupo de capacidad de EC2 determinado (el número total de slots del grupo menos el número de instancias en ejecución)
- $serverSlots_{max}$ es el número máximo de slots que cualquier servidor aporta a un grupo de capacidad de EC2 determinado
- M es el número de servidores que pueden fallar y, aun así, permitir que los servidores restantes alojen las instancias en ejecución

Ejemplo: un Outpost tiene tres servidores que aportan ranuras a una reserva de capacidad. `m5.2xlarge` El primero aporta 4 slots, el segundo aporta 3 y el tercero aporta 2. El grupo de `m5.2xlarge` instancias del Outpost tiene una capacidad total de 9 ranuras (4 + 3 + 2). El Outpost tiene 4 instancias en ejecución`m5.2xlarge`. ¿Cuántos servidores pueden fallar y, aun así, permitir que los servidores restantes alojen las instancias en ejecución?

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

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

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

Respuesta: Cualquiera de los servidores puede fallar y permitir el mantenimiento de las instancias en ejecución en los servidores restantes.

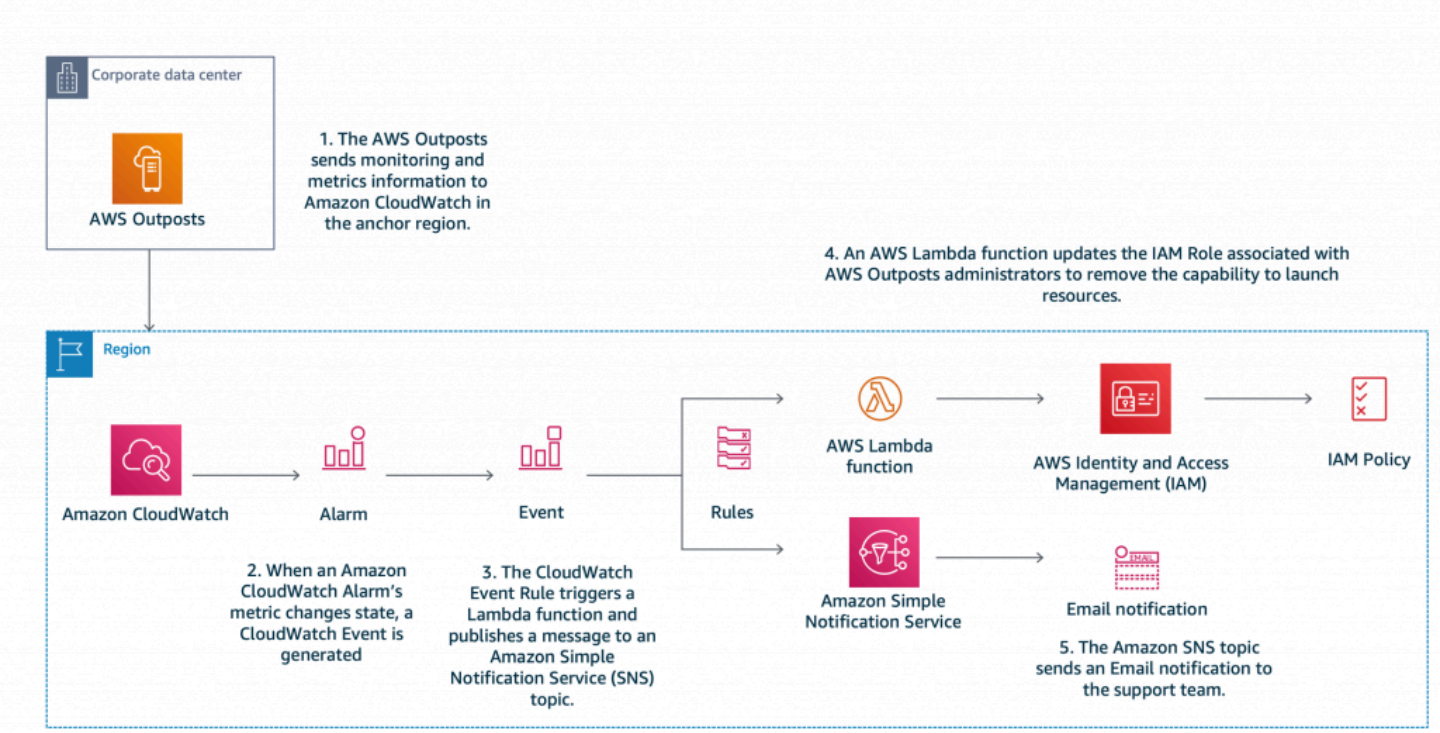
Prácticas recomendadas para la planificación de la capacidad informática

- Dimensione su capacidad informática para proporcionar redundancia N+M para cada grupo de capacidad de EC2 de Outposts.
  - Implemente servidores N+M para servidores con configuraciones homogéneas de slots, o bien heterogéneas e idénticas.
  - Calcule la disponibilidad N+M para cada grupo de capacidad de EC2 y asegúrese de que cada grupo satisfaga los requisitos de disponibilidad.

Administración de la capacidad

Puede supervisar el uso del grupo de instancias EC2 de Outpost en las métricas de Amazon AWS Management Console CloudWatch y a través de ellas. Póngase en contacto con Enterprise Support para recuperar o cambiar los diseños de slots de las implementaciones de Outposts.

Utilice los mismos mecanismos de recuperación automática de instancias y Auto Scaling de EC2 para recuperar o reemplazar las instancias afectadas por fallos del servidor y eventos de mantenimiento. Se debe supervisar y administrar la capacidad de Outposts para garantizar que siempre haya suficiente capacidad de reserva disponible para adaptarse a los errores del servidor. La publicación Managing AWS Outposts your capacity using Amazon CloudWatch and AWS Lambda blog contiene un tutorial práctico en el que se muestra cómo combinar AWS CloudWatch y gestionar la capacidad de Outpost AWS Lambda para mantener la disponibilidad de las instancias.



Administrar AWS Outposts la capacidad con Amazon CloudWatch y AWS Lambda

Prácticas recomendadas para la gestión de la capacidad informática

- Configure las instancias de EC2 en grupos de escalado automático o utilice la recuperación automática de instancias para reiniciar las instancias con errores.
- Automatice la supervisión de la capacidad de las implementaciones de Outposts y configure las notificaciones y (opcionalmente) las respuestas automatizadas para las alarmas de capacidad.

Ubicación de instancias

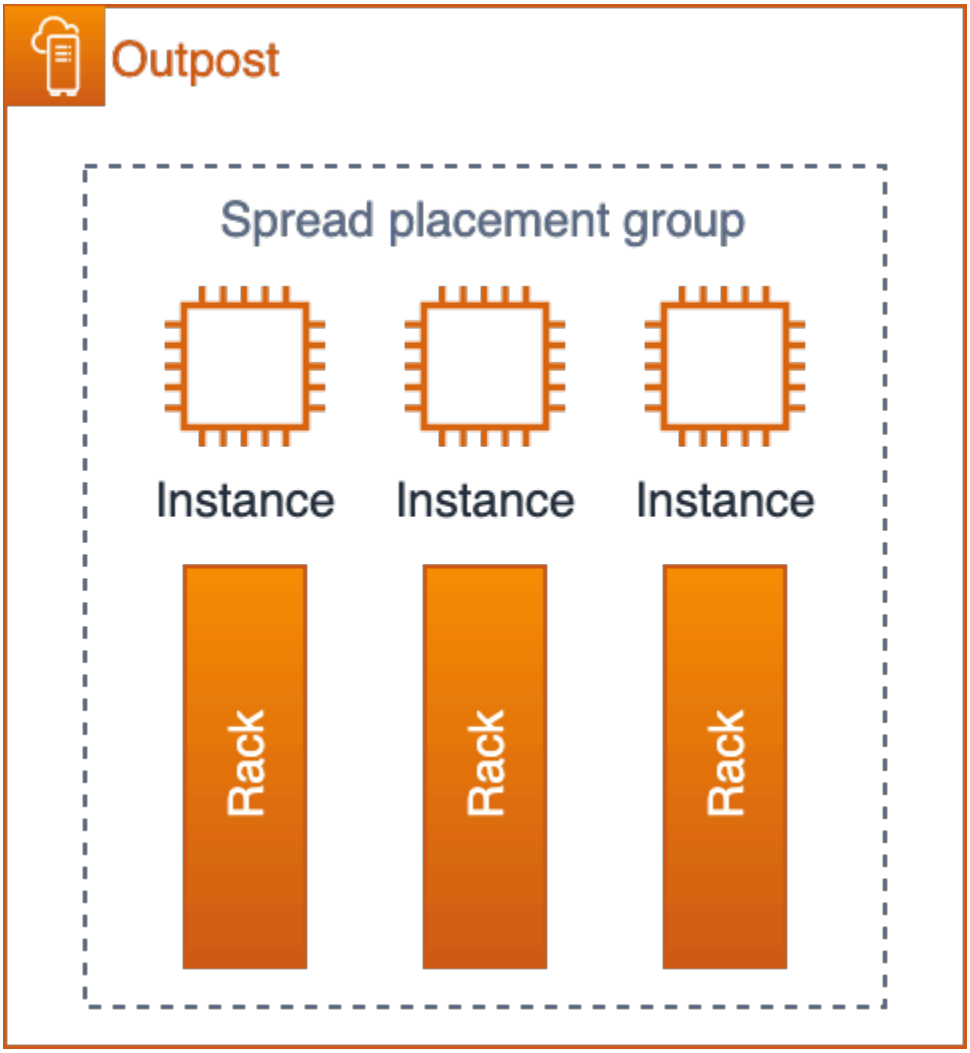
Las implementaciones de Outposts tienen un número finito de servidores informáticos. Si una aplicación implementa varias instancias relacionadas en Outposts, sin configuración adicional, las instancias pueden implementarse en el mismo servidor o en servidores del mismo bastidor. En la actualidad, existen tres mecanismos para distribuir las instancias a fin de mitigar el riesgo de ejecutar instancias relacionadas en la misma infraestructura:

Implementación de varias instancias de Outposts: de forma similar a una estrategia con múltiples zonas de disponibilidad en la región, se pueden implementar varias instancias de Outposts en centros de datos independientes, así como recursos de aplicaciones para instancias específicas de Outposts. Esto permite ejecutar instancias en la implementación de Outposts deseada (un

conjunto lógico de bastidores). Se puede emplear una estrategia de múltiples instancias de Outposts para protegerse contra los modos de error del bastidor y del centro de datos y, si las instancias de Outposts están ancladas a AZ o regiones independientes, también pueden brindar protección frente a los modos de error de las AZ o la región. Para obtener más información acerca de las arquitecturas con múltiples implementaciones de Outposts, consulte la publicación Modos de error más extensos.

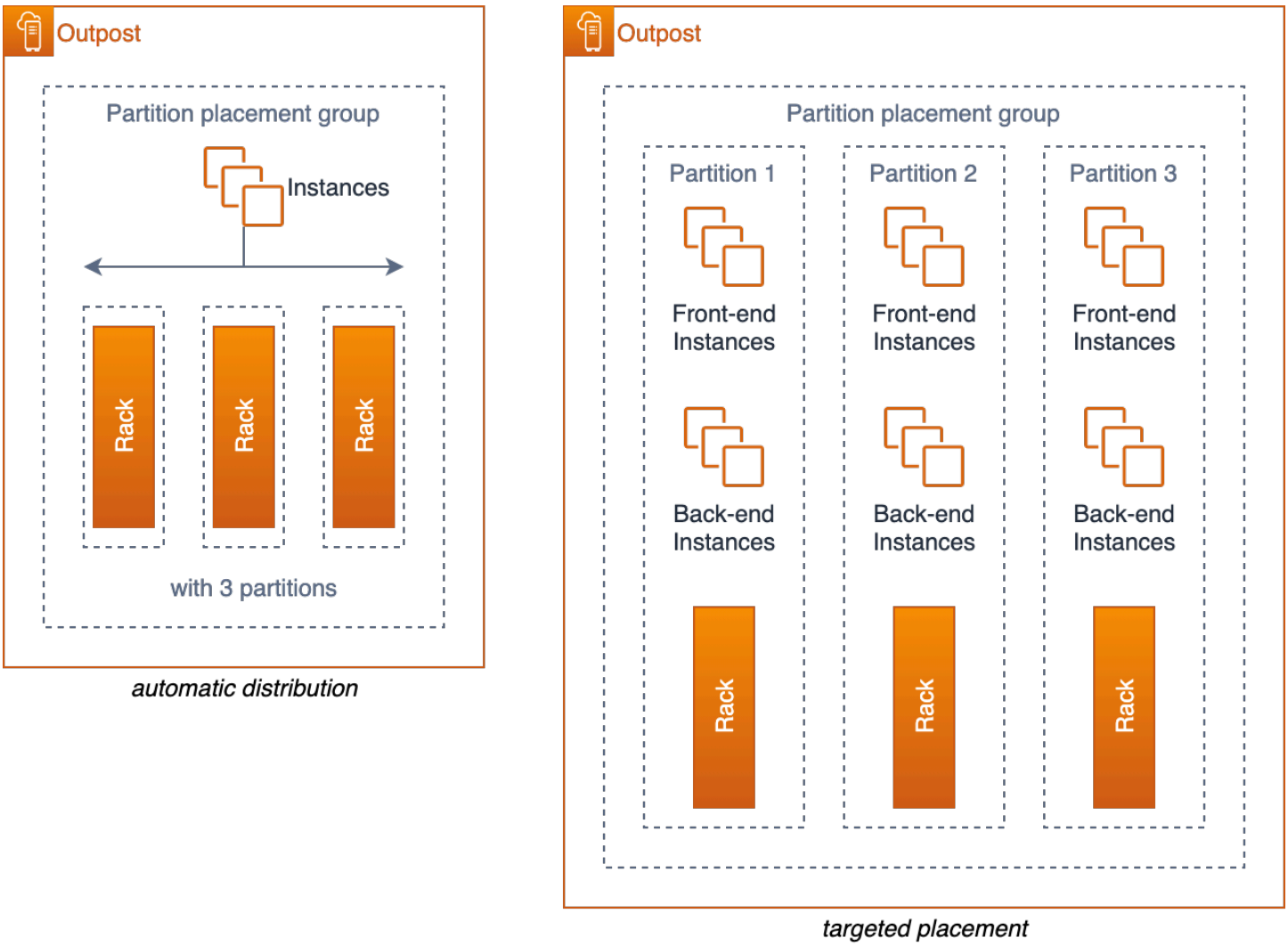
Grupos de ubicación de Amazon EC2 en Outposts (ubicación de instancias con varios bastidores de implementaciones únicas de Outposts): permiten utilizar las estrategias de clúster, distribución y partición para influir en la ubicación. Las estrategias de distribución y partición en torno a la ubicación permiten distribuir las instancias entre los bastidores de una implementación de Outposts con varios bastidores.

Un grupo con ubicación distribuida proporciona una forma sencilla de distribuir las instancias individuales entre los bastidores para reducir la posibilidad de que se produzcan errores correlacionados. Solo se puede implementar en el grupo el número de instancias que iguale el número de bastidores de Outposts.



Grupo con ubicación distribuida de EC2 en una implementación de Outposts con tres bastidores

También se pueden distribuir las instancias en varios bastidores con grupos con ubicación en particiones. La distribución automática se utiliza para distribuir las instancias entre las particiones del grupo o implementar las instancias en las particiones de destino seleccionadas. La implementación de instancias en las particiones de destino permite implementar los recursos seleccionados en el mismo bastidor y, al mismo tiempo, distribuir otros recursos entre todos los bastidores. Por ejemplo, si el usuario dispone de una instancia lógica de Outposts con tres bastidores, crear un grupo con ubicación en particiones con tres particiones le va a permitir distribuir los recursos entre los bastidores.



Grupo con ubicación en particiones de EC2 en una implementación de Outposts con tres bastidores

Configuración creativa de slots para servidores: si el usuario cuenta con una implementación de Outposts de un solo bastidor o si el servicio que utiliza en Outposts no admite grupos de ubicación, es posible que pueda utilizar una configuración de slots creativa para que las instancias no se implementen en el mismo servidor físico. Si las instancias relacionadas tienen el mismo tamaño de instancia de EC2, es posible configurar los slots de los servidores para limitar la cantidad de slots de ese tamaño configuradas en cada servidor, distribuyendo los slots entre los distintos servidores. La configuración de slots de los servidores limitará el número de instancias (de ese tamaño) que se pueden ejecutar en un único servidor.

Un ejemplo es el diseño de configuración de slots que se ha mostrado anteriormente en la figura 13. Si su aplicación necesitara implementar tres `m5.4xlarge` instancias en el Outpost configurado con este diseño de ranuras, EC2 colocaría cada instancia en un servidor independiente y no habría

ninguna posibilidad de que estas instancias pudieran ejecutarse en el mismo servidor, siempre y cuando la configuración de asignación de ranuras no cambie para abrir `m5.4xlarge` ranuras adicionales en los servidores.

Prácticas recomendadas para la ubicación de instancias informáticas

- Utilice los grupos de ubicación de Amazon EC2 en Outposts para controlar la ubicación de las instancias en los bastidores de una implementación única de Outposts.
- En lugar de pedir una sola instancia de Outposts con un único bastidor de Outposts de tamaño mediano o grande, el usuario debe plantearse la división de la capacidad en dos bastidores pequeños o medianos a fin de poder aprovechar la capacidad de los grupos de ubicación de EC2 para distribuir las instancias entre los distintos bastidores.

Almacenamiento

El servicio de almacenamiento en AWS Outposts rack ofrece tres tipos de almacenamiento:

- Almacenamiento de instancias en tipos de instancias de EC2 compatibles
- Volúmenes gp2 de Amazon Elastic Block Store (EBS) para el almacenamiento de bloques persistentes
- Amazon Simple Storage Service en Outposts (S3 en Outposts) para el almacenamiento de objetos locales

El almacenamiento de instancias se proporciona en servidores compatibles (`C5d`, `M5d`, `R5d`, `G4dn` y `I3en`). Al igual que en la región, los datos de un almacén de instancias solo se conservan durante la vida útil (ejecución) de la instancia.

Los volúmenes EBS de Outposts y el almacenamiento de objetos S3 en Outposts se proporcionan como parte de los servicios administrados de bastidores de AWS Outposts . Los clientes son responsables de la administración de la capacidad de los grupos de almacenamiento de Outposts. Los clientes especifican sus requisitos de almacenamiento para EBS y S3 al solicitar un Outpost. AWS configura el Outpost con el número de servidores de almacenamiento necesarios para proporcionar la capacidad de almacenamiento solicitada. AWS es responsable de la disponibilidad de los servicios de almacenamiento de EBS y S3 en Outposts. Se han aprovisionado suficientes servidores de almacenamiento para proporcionar servicios de almacenamiento de alta disponibilidad a Outposts. La pérdida de un único servidor de almacenamiento no debería interrumpir los servicios ni ocasionar la pérdida de datos.

Puedes usar las CloudWatch métricas AWS Management Console y para monitorear la utilización de la capacidad de Outpost, EBS y S3 sobre la utilización de la capacidad de Outposts.

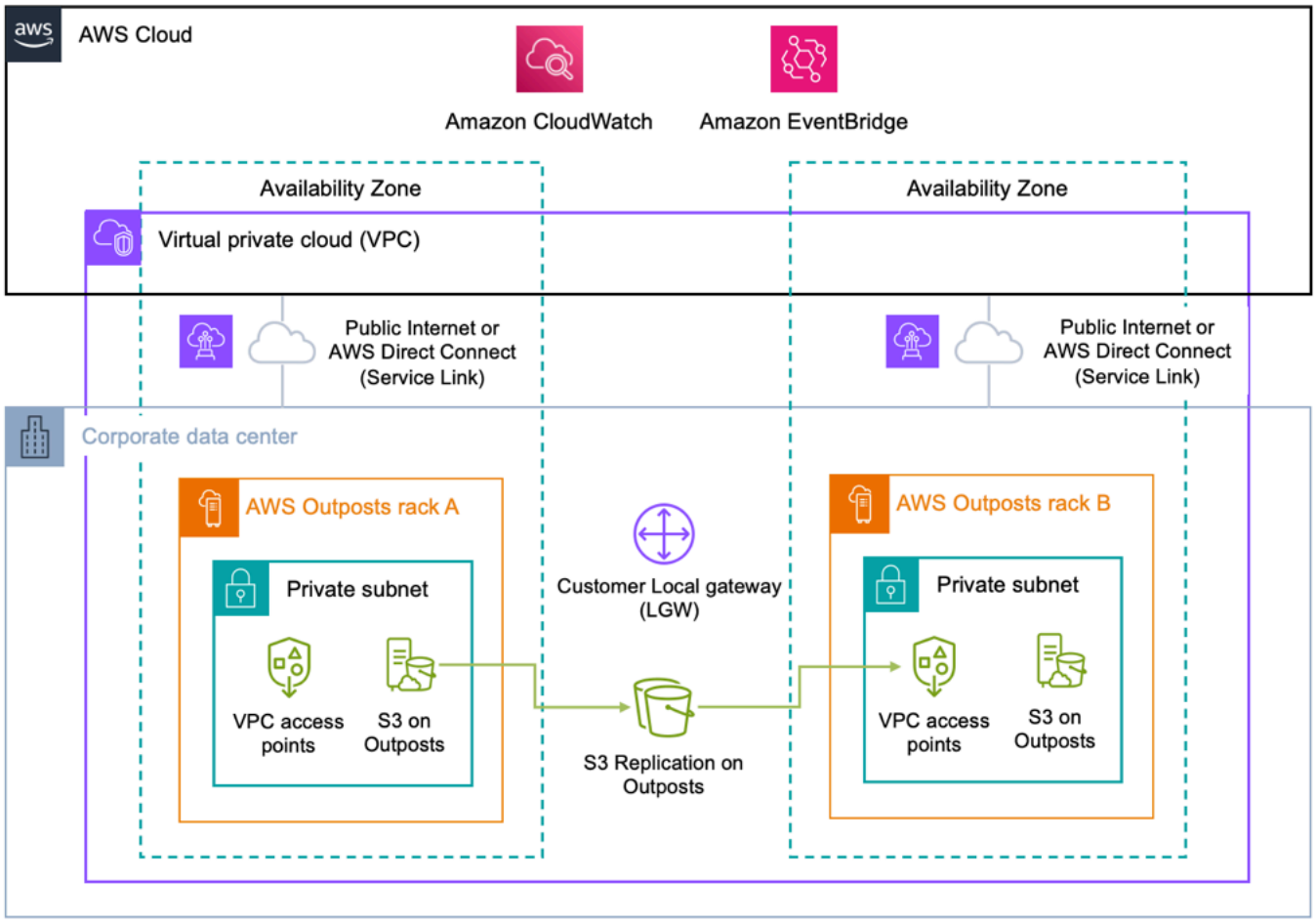
Protección de datos

Para los volúmenes de EBS: AWS Outposts rack admite instantáneas de volúmenes de EBS para proporcionar un mecanismo de protección de datos simple y seguro que proteja los datos de almacenamiento en bloque. Las instantáneas son copias de seguridad point-in-time incrementales de sus volúmenes de EBS. De forma predeterminada, las instantáneas de los volúmenes de Amazon EBS de Outposts se almacenan en Amazon S3, en la región correspondiente. Si la capacidad de Outposts se ha configurado con S3, se pueden usar instantáneas locales de EBS en Outposts para almacenar las instantáneas localmente en la implementación de Outposts utilizando el almacenamiento de S3 en Outposts.

Para los buckets de S3 en Outposts (casos de uso de residencia de datos):

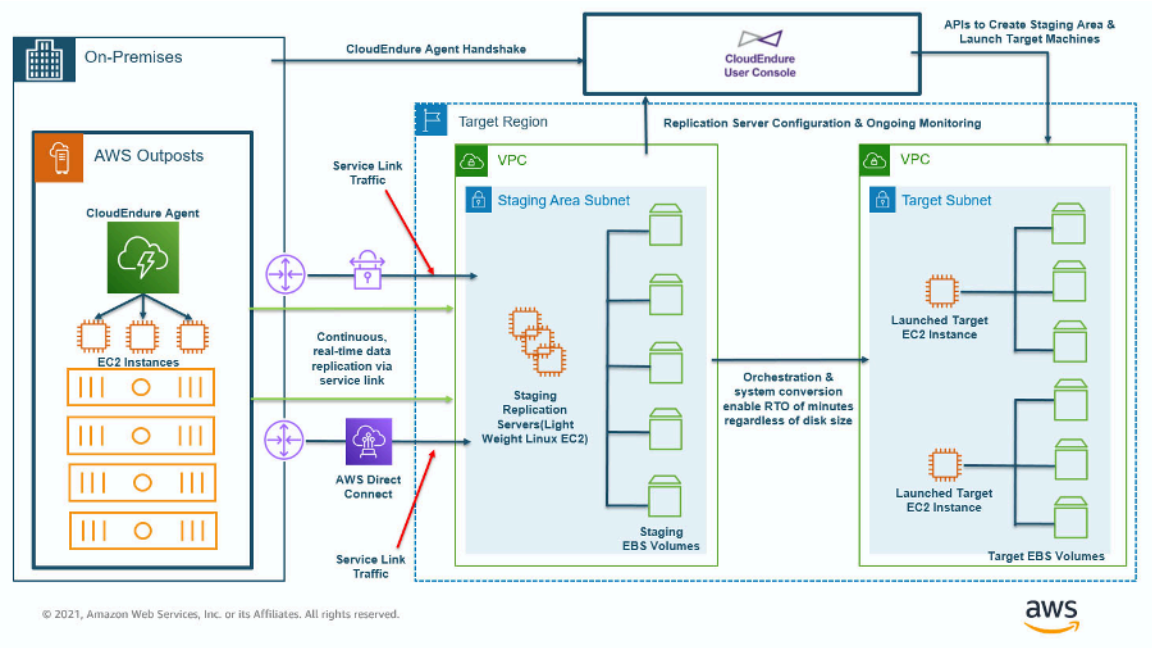
- El control de versiones de S3 en Outposts se puede utilizar para guardar todos los cambios y el historial de los objetos. Cuando está habilitado, el control de versiones de S3 guarda diversas copias de un objeto en el mismo bucket. Puede utilizar el control de versiones de S3 para conservar, recuperar y restaurar todas las versiones de los objetos almacenados en su bucket de Outposts. EL control de versiones de S3 ayuda a recuperarse de acciones no deseadas del usuario y de errores de la aplicación.
- La replicación de S3 en Outposts se puede utilizar para crear y configurar reglas de replicación que repliquen automáticamente los objetos de S3 en otra instancia de Outposts o en otro bucket de la misma instancia de Outposts. Durante la replicación, los objetos de S3 en Outposts se envían a través de la puerta de enlace local (LGW) del cliente y los objetos no regresan a Región de AWS. La replicación de S3 en Outposts representa una forma fácil y flexible de replicar automáticamente los datos dentro de un perímetro de datos específico para abordar los requisitos de redundancia y conformidad de los datos.

  La replicación de S3 en Outposts también proporciona métricas detalladas y notificaciones para supervisar el estado de la replicación de los objetos. Puedes monitorizar el progreso de la replicación mediante el seguimiento de los bytes pendientes, las operaciones pendientes y la latencia de replicación entre los depósitos de Outposts de origen y destino mediante Amazon. CloudWatch También puede configurar EventBridge las reglas de Amazon para recibir eventos de error de replicación a fin de diagnosticar y corregir rápidamente los problemas de configuración.



Para grupos de S3 on Outposts (casos de uso no relacionados con la residencia de datos) Regiones de AWS: puedes AWS DataSyncutilizarlos para automatizar las transferencias de datos de S3 on Outposts entre tu Outpost y la región. DataSync te permite elegir qué transferir, cuándo transferir y cuánto ancho de banda usar. Hacer copias de seguridad en las instalaciones de los buckets de S3 en Outposts a buckets de S3 en Región de AWS permite aprovechar el 99,999999999 % (11 nueves) de durabilidad de los datos y los niveles de almacenamiento adicionales (Standard, Infrequent Access y Glacier) para optimizar los costos disponibles con el servicio S3 regional.

Replicación de instancias: se puede utilizar CloudEndurepara replicar instancias individuales de sistemas locales a un puesto de avanzada, de un puesto de avanzada a la región, de la región a un puesto de avanzada o de un puesto de avanzada a otro. La entrada del CloudEndure blog Architecting for DR on AWS Outposts with describe cada uno de estos escenarios y cómo diseñar una solución con ellos. CloudEndure



Recuperación de desastres desde una implementación de Outposts a la región

El uso de AWS Outposts rack como CloudEndure destino (objetivo de replicación) requiere S3 on Outposts Storage.

Prácticas recomendadas para la protección de datos

- Utilice las instantáneas de EBS para crear point-in-time copias de seguridad de los volúmenes de almacenamiento en bloque en Amazon S3 en la región o en S3 en Outposts.
- Use el control de versiones de objetos de S3 en Outposts para mantener múltiples versiones y el historial de objetos.
- Emplee la replicación de S3 en Outposts para replicar automáticamente los datos de los objetos en otra implementación de Outposts.
- Para los casos de uso no relacionados con la residencia de datos, AWS DataSync úselo para hacer copias de seguridad de los objetos almacenados en S3 en Outpost en Amazon S3 de la región.
- Úselo CloudEndure para replicar instancias entre sistemas locales, Outposts lógicos y la región.

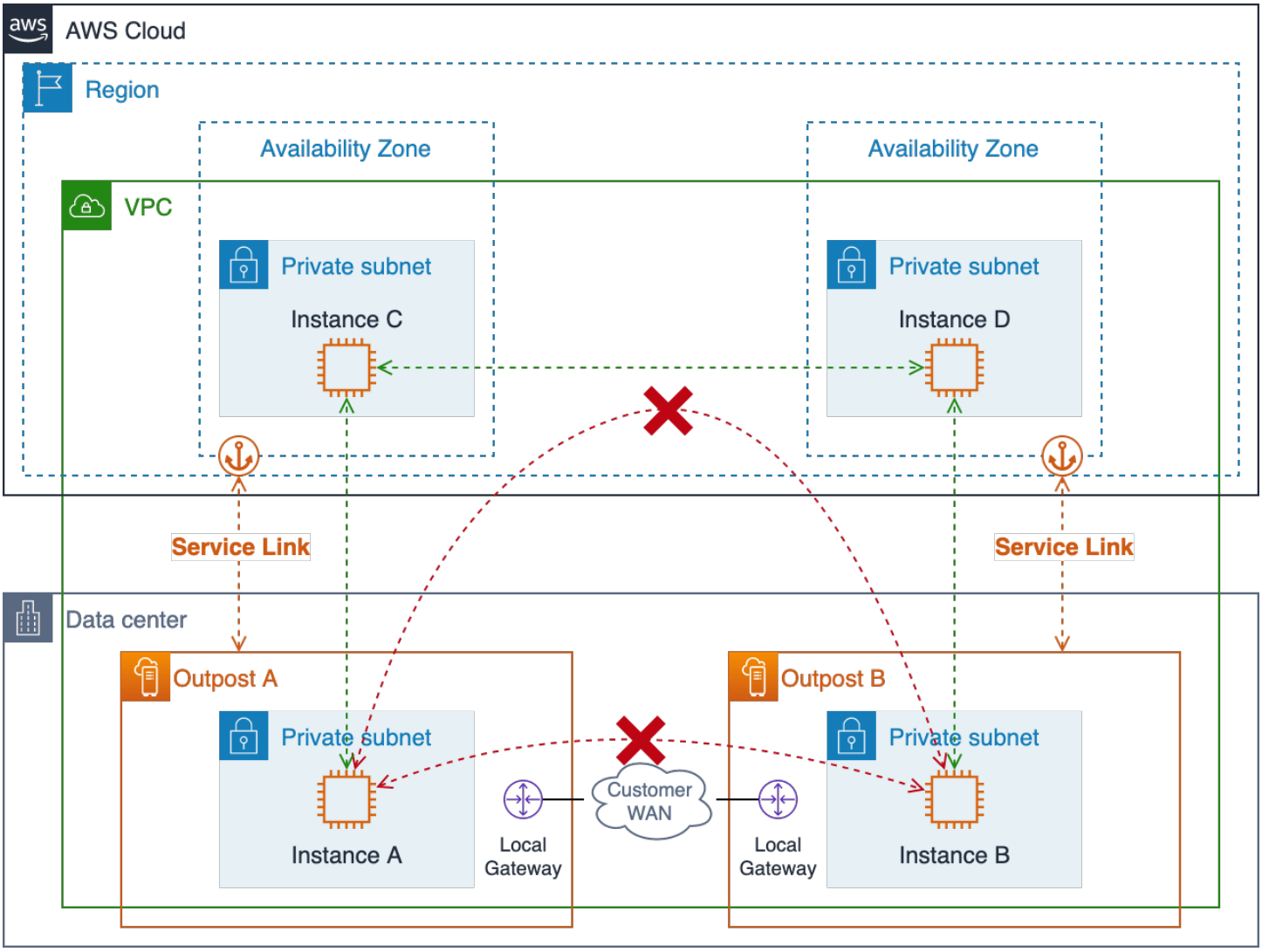
Modos de error más extensos

Para diseñar arquitecturas de alta disponibilidad que mitiguen los modos de error más extensos, como errores de bastidor, de centro de datos, de zonas de disponibilidad (AZ) o de regiones, es

necesario implementar varias instancias de Outposts con suficiente capacidad de infraestructura en centros de datos separados con conectividad WAN y alimentación independientes. Las implementaciones de Outposts se anclan a distintas zonas de disponibilidad (AZ) dentro de una Región de AWS o varias regiones. También debe proporcionar una site-to-site conectividad flexible y suficiente entre las ubicaciones para admitir la replicación de datos sincrónica o asíncrona y la redirección del tráfico de la carga de trabajo. Según la arquitectura de la aplicación, se puede utilizar el DNS de Amazon Route 53 disponible globalmente y los servicios de Elastic Load Balancing disponibles regionalmente para dirigir el tráfico a la ubicación deseada y automatizar la redirección del tráfico a las ubicaciones restantes en caso de que se produzcan errores a gran escala.

Existen limitaciones de red que hay que tener en cuenta al diseñar e implementar cargas de trabajo de aplicaciones en varias instancias de Outposts. Los recursos de dos instancias independientes de Outposts no pueden comunicarse entre sí mediante el tránsito de tráfico por la región. Los recursos de dos instancias independientes de Outposts implementadas en la misma VPC no pueden comunicarse entre sí a través de la red del cliente. Los recursos de dos instancias independientes de Outposts implementadas en distintas VPC sí pueden comunicarse entre sí a través de la red del cliente.

Las dos figuras siguientes ilustran las rutas de red bloqueadas y correctas.

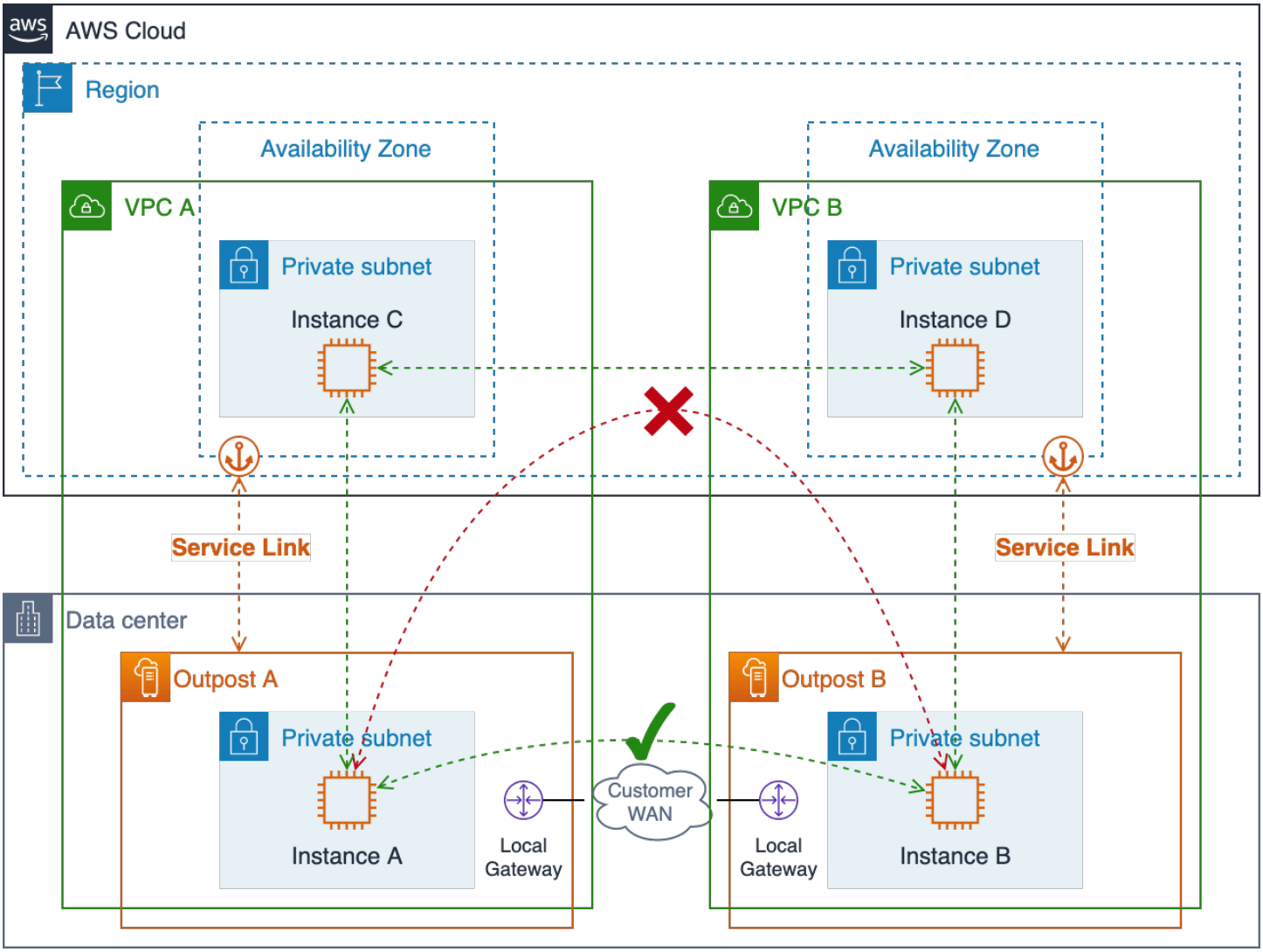


Rutas de red de varias implementaciones de Outposts en una sola VPC

El tráfico entre dos implementaciones de Outposts que transita por la región se encuentra bloqueado, ya que se trata de un antipatrón. Este tráfico generaría gastos de salida en ambas direcciones y es probable que la latencia fuera mucho mayor que la de simplemente enrutar el tráfico a través de la WAN del cliente.

Los recursos de varias implementaciones de Outposts en la misma VPC no pueden comunicarse entre sí. El tráfico entre implementaciones de Outposts en la misma VPC siempre seguirá la ruta CIDR de la VPC local a través de la región en la que se bloqueará.

Se deberían usar VPC independientes para implementar recursos en varias implementaciones de Outposts y poder así enrutar el tráfico de una implementación de Outposts a otra a través de las redes locales en las instalaciones y las redes WAN del usuario.



Rutas de red de varias implementaciones de Outposts en varias VPC

Prácticas recomendadas para protegerse frente a modos de error más extensos

- Implemente varias instancias de Outposts ancladas a varias AZ y regiones.
- Use VPC independientes para cada instancia de Outposts en una implementación con varias.

Conclusión

Con AWS Outposts rack, puede crear, gestionar y escalar aplicaciones locales de alta disponibilidad mediante AWS herramientas y servicios conocidos, como Amazon EC2, Amazon EBS, Amazon S3 on Outposts, Amazon ECS, Amazon EKS y Amazon RDS. Las cargas de trabajo pueden ejecutarse de forma local, servir a clientes, acceder a las aplicaciones y los sistemas de las redes en las instalaciones y acceder al conjunto completo de servicios de Región de AWS. Los bastidores de Outposts son ideales para cargas de trabajo que requieren acceso de baja latencia a sistemas en las instalaciones, procesamiento de datos local, residencia de datos y migración de aplicaciones con interdependencias de sistemas locales.

Si proporciona una implementación de Outpost con la energía, el espacio y la refrigeración adecuados y conexiones flexibles Región de AWS, puede crear servicios de centro de datos únicos de alta disponibilidad. Para obtener niveles más altos de disponibilidad y resiliencia, se pueden implementar varias instancias de Outposts y distribuir las aplicaciones más allá de límites lógicos y geográficos.

El rack Outposts elimina el pesado trabajo indiferenciado de crear grupos de redes de aplicaciones, almacenamiento y cómputo locales y le permite extender el alcance de la infraestructura AWS global a sus centros de datos e instalaciones de ubicación conjunta. Los usuarios ya pueden dedicar su tiempo y energía a modernizar sus aplicaciones, agilizar las implementaciones y hacer que los servicios de TI repercutan más y mejor en la empresa.

Colaboradores

Los colaboradores de este documento son:

- Mallory Gershenfeld, S3 en Outposts, Amazon Web Services
- Chris Lunsford, arquitecto sénior especializado en soluciones AWS Outposts, Amazon Web Services
- Rohan Mathews, arquitecto principal de Amazon AWS Outposts Web Services

Historial del documento

Para recibir notificaciones sobre las actualizaciones de este documento técnico, suscríbase a la fuente RSS.

Cambio	Descripción	Fecha
Actualización menor	Se agregó una guía adicional de asignación de fechas en la planificación de la capacidad.	9 de febrero de 2024
Actualización menor	Se ha actualizado para reflejar el lanzamiento de nuevas características desde su publicación inicial.	19 de julio de 2023
Actualización menor	Se han actualizado las prácticas recomendadas para conexiones de redes de alta disponibilidad.	29 de junio de 2023
Publicación inicial	Documento técnico publicado por primera vez.	12 de agosto de 2021

> **Note**
>
> Para suscribirse a las actualizaciones de RSS, debe tener un complemento de RSS habilitado para el navegador que esté utilizando.

Avisos

Es responsabilidad de los clientes realizar su propia evaluación independiente de la información que contiene este documento. Este documento: (a) tiene únicamente fines informativos, (b) representa las ofertas y prácticas de AWS productos actuales, que están sujetas a cambios sin previo aviso, y (c) no crea ningún compromiso ni garantía por parte de AWS sus filiales, proveedores o licenciantes. AWS los productos o servicios se proporcionan «tal cual» sin garantías, representaciones o condiciones de ningún tipo, ya sean expresas o implícitas. Las responsabilidades y obligaciones de AWS sus clientes están reguladas por AWS acuerdos, y este documento no forma parte de ningún acuerdo entre sus clientes AWS y sus clientes ni lo modifica.

AWS Glosario

Para obtener la AWS terminología más reciente, consulte el AWS glosario de la Glosario de AWS Referencia.

Las traducciones son generadas a través de traducción automática. En caso de conflicto entre la traducción y la version original de inglés, prevalecerá la version en inglés.