Documento técnico de AWS

Conectividad híbrida

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

Conectividad híbrida: Documento técnico de AWS

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

Las marcas comerciales y la imagen comercial de Amazon no se pueden utilizar en relación con ningún producto o servicio que no sea de Amazon, de ninguna manera que pueda causar confusión entre los clientes y que menosprecie o desacredite a Amazon. Todas las demás marcas comerciales que no son propiedad de Amazon son propiedad de sus respectivos propietarios, que pueden o no estar afiliados, relacionados o patrocinados por Amazon.

Table of Contents

Resumen e introducción i
Introducción 1
¿Usa Well-Architected? 2
Componentes de conectividad híbrida de AWS 3
Conexiones de red híbrida 3
AWS Direct Connect 3
Site-to-Site VPN 5
Transit Gateway Connect 6
Servicios de conectividad híbrida de AWS 6
Consideraciones sobre el tipo de conectividad híbrida y su diseño 8
Selección del tipo de conectividad 9
Momento de implementación 10
Seguridad 12
Acuerdo de nivel de servicios 13
Desempeño 15
Costo 18
Selección de diseño de conectividad 22
Escalabilidad 22
Modelos de conectividad 23
Fiabilidad 36
VPN y SD-WAN administradas por el cliente 44
Caso de uso de Example Corp. Automotive 47
Arquitectura seleccionada 54
Conclusión 56
Colaboradores 57
Documentación adicional 58
Revisiones del documento 59
Avisos 60
Glosario de AWS 61

Conectividad híbrida

Fecha de publicación: 6 de julio de 2023 (Revisiones del documento)

Muchas organizaciones necesitan conectar sus centros de datos en las instalaciones, los sitios remotos y la nube. Una red híbrida conecta estos distintos entornos. En este documento técnico se describen los componentes de AWS y los requisitos clave que deben tenerse en cuenta a la hora de decidir qué modelo de conectividad híbrida le resulta más adecuado. Para ayudarlo a determinar la mejor solución para sus requisitos técnicos y empresariales, le ofrecemos árboles de decisiones que lo guiarán a lo largo del proceso lógico de selección.

Introducción

Una organización moderna utiliza una amplia gama de recursos de TI. En el pasado, era habitual alojar estos recursos en un centro de datos en las instalaciones o en una instalación de coubicación. Con la creciente adopción de la computación en la nube, las organizaciones entregan y consumen recursos de TI de proveedores de servicios en la nube a través de una conexión de red. Las organizaciones pueden optar por migrar algunos de sus recursos de TI existentes, o todos, a la nube. En cualquier caso, se necesita una red común para conectar los recursos en las instalaciones y en la nube. La coexistencia de recursos en las instalaciones y en la nube se denomina nube híbrida y la red común que los conecta se denomina red híbrida. Incluso si su organización mantiene todos sus recursos de TI en la nube, es posible que siga necesitando conectividad híbrida con sitios remotos.

Hay varios modelos de conectividad entre los que elegir. Aunque disponer de opciones agrega flexibilidad, la selección de la opción óptima requiere un análisis de los requisitos empresariales y técnicos, así como la eliminación de las opciones que no sean adecuadas. Puede agrupar los requisitos en función de consideraciones como la seguridad, el tiempo de implementación, el rendimiento, la fiabilidad, el modelo de comunicación, la escalabilidad, etc. Los arquitectos de redes y de la nube, una vez que han recopilado, analizado y considerado detenidamente los requisitos, pueden identificar los componentes básicos y las soluciones de red híbrida de AWS aplicables. Para identificar y seleccionar el modelo o los modelos óptimos, los arquitectos deben comprender las ventajas y desventajas de cada modelo. También existen limitaciones técnicas que pueden provocar la exclusión de un modelo adecuado.

Para simplificar el proceso de selección, este documento técnico lo guía por cada consideración clave en un orden lógico. En cada consideración, hay preguntas que se utilizan para recopilar los requisitos. Se identifica el impacto de cada decisión de diseño, junto con las soluciones potenciales.

En el documento técnico se presentan árboles de decisiones para algunas de las consideraciones como método para ayudar en el proceso de toma de decisiones, eliminar opciones y comprender las consecuencias de cada decisión. Concluye con un escenario que trata un caso de uso híbrido, mediante la aplicación de la selección y el diseño del modelo de conectividad de extremo a extremo. Puede utilizar este ejemplo para ver cómo ejecutar los procesos expuestos en este documento técnico en un ejemplo práctico.

El objetivo de este documento técnico es ayudarlo a seleccionar y diseñar un modelo óptimo de conectividad híbrida. Este documento técnico está estructurado de la siguiente manera:

- Componentes básicos de conectividad híbrida: información general de los servicios de AWS utilizados para la conectividad híbrida.
- Consideraciones sobre la selección y el diseño de la conectividad: una definición de cada modelo de conectividad, cómo cada uno afecta la decisión de diseño, preguntas de identificación de requisitos, soluciones y árboles de decisiones.
- Un caso de uso de cliente: un ejemplo de cómo utilizar en la práctica las consideraciones y los árboles de decisiones.

¿Usa Well-Architected?

El Marco de AWS Well-Architected lo ayuda a comprender los pros y los contras 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 una serie de preguntas para cada pilar.

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

Componentes de conectividad híbrida de AWS

Una arquitectura de conectividad de red híbrida consta de tres componentes básicos:

- Conexiones de red híbrida: los tipos de conexión entre los servicios de conectividad de AWS y los dispositivos de puerta de enlace de cliente en las instalaciones.
- Servicios de conectividad híbrida de AWS: los servicios de AWS que proporcionan conectividad y enrutamiento entre la infraestructura del cliente y AWS.
- Dispositivo de puerta de enlace de cliente en las instalaciones: el dispositivo en la red existente del cliente que es el punto de conexión en las instalaciones para la conexión a la red híbrida. Los distintos tipos de conexión tienen diferentes requisitos técnicos para estos dispositivos, que se analizan en las siguientes secciones.

Conexiones de red híbrida

Existen varias formas de conectar sus equipos en las instalaciones y AWS. El presente documento técnico se centra en cómo pueden combinarse estas diferentes vías en arquitecturas globales. No obstante, se ofrece una breve descripción general de las diferentes opciones (AWS Direct Connect, red privada virtual de sitio a sitio y Transit Gateway Connect).

AWS Direct Connect

AWS Direct Connect es un servicio que establece una conexión de red dedicada entre las instalaciones locales y AWS. Para obtener más información, consulte AWS Direct Connect.

Existen dos tipos de conexiones de AWS Direct Connect: dedicadas y alojadas. Una conexión dedicada es un enlace directo entre un dispositivo de AWS y su dispositivo en las instalaciones, mientras que una conexión alojada es una conexión para la que brinda soporte un socio de AWS que puede gestionar los detalles de la conexión por usted. Consulte Conexiones de AWS Direct Connect para obtener más información.

Una conexión de Direct Connect utiliza interfaces virtuales (VIF) para aislar los distintos flujos de tráfico. Varias VIF pueden utilizar el mismo enlace de Direct Connect, separadas por etiquetas de VLAN (802.1q). Hay tres tipos de VIF que proporcionan conectividad a la red de AWS. Consulte Interfaces virtuales de AWS Direct Connect para obtener más detalles. Los tres tipos son:

- VIF privada: una VIF privada es una conexión privada entre el dispositivo y los recursos en AWS. Terminan en AWS, ya sea en una puerta de enlace privada virtual (VGW) directamente (que admite una sola VPC) o a través de una puerta de enlace de Direct Connect que, después, se conecta a varias VGW.
- VIF pública: una VIF pública permite la conectividad con cualquier recurso de AWS público, como S3, DynamoDB e intervalos de IP EC2 públicos. Aunque una VIF pública no tiene acceso directo a Internet, cualquier recurso público de Amazon puede llegar a ella (incluidas las instancias de EC2 públicas de otros clientes), lo que los clientes deben tener en cuenta al planificar la seguridad.
- VIF de tránsito: una VIF de tránsito es una conexión privada entre su dispositivo y AWS Transit Gateway, a través de una puerta de enlace de Direct Connect. Las VIF de tránsito ahora se admiten en enlaces con velocidades inferiores a 1 Gbps. Consulte el anuncio de lanzamiento para obtener más detalles.

> **Note**
>
> La interfaz virtual alojada (VIF alojada) es un tipo de VIF privada que se asigna a una Cuenta de AWS distinta de la Cuenta de AWS propietaria de la conexión de AWS Direct Connect (que puede ser un socio de AWS Direct Connect). AWS ya no permite que nuevos socios ofrezcan este modelo. Para obtener más información, consulte Creación de una interfaz virtual alojada.

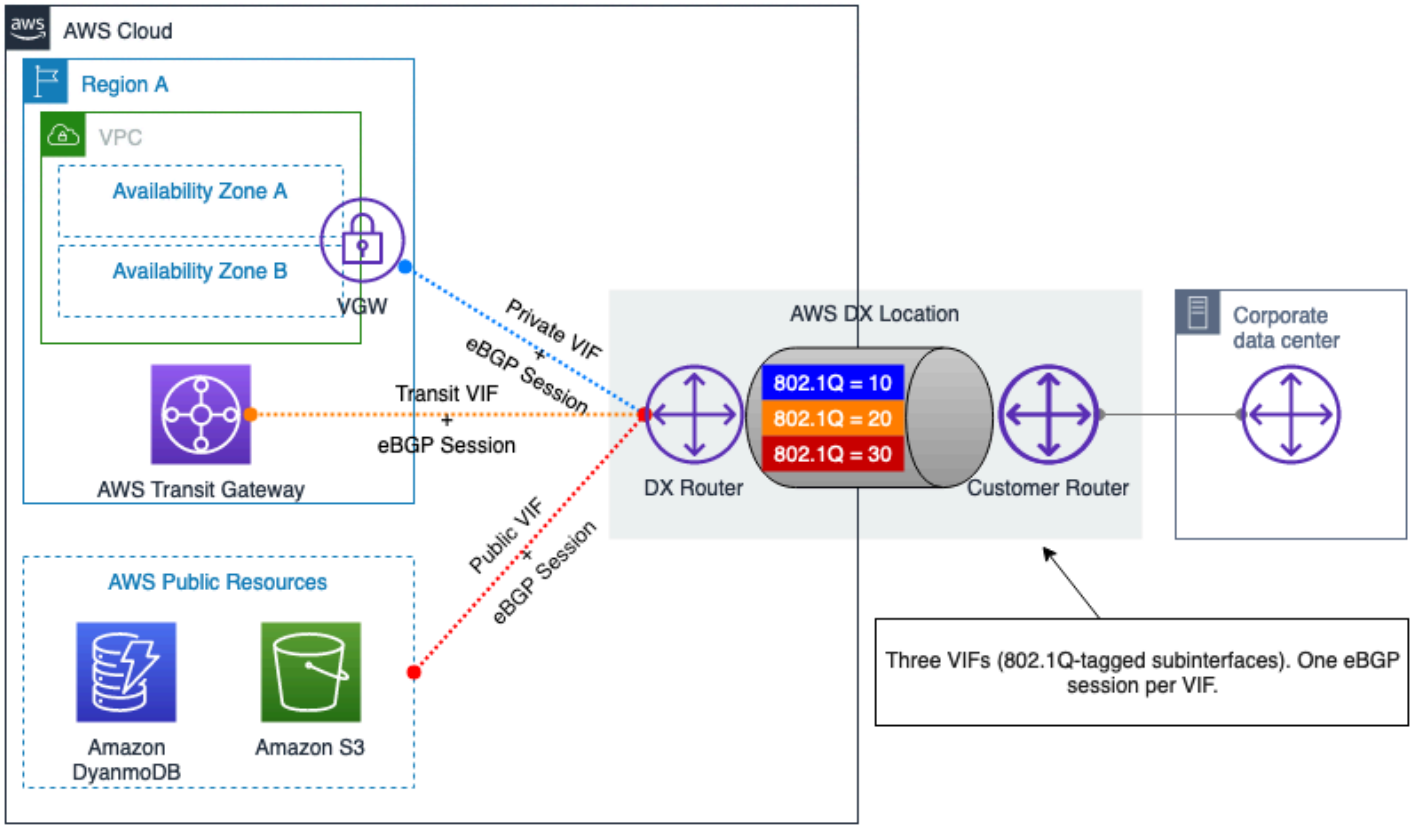

Figura 1. VIF privadas y públicas de AWS Direct Connect

Red privada virtual (VPN) de sitio a sitio

Una VPN de sitio a sitio permite que dos redes se comuniquen de forma segura y puede utilizarse a través de un medio de transporte que no sea de confianza, como Internet. Los clientes pueden establecer conexiones de VPN entre los sitios en las instalaciones y Amazon Virtual Private Cloud (Amazon VPC) mediante dos opciones:

- VPN de sitio a sitio administrada por AWS (AWS S2S VPN): se trata de un servicio VPN totalmente administrado y de alta disponibilidad, que utiliza IPSec. Consulte ¿Qué es AWS Site-to-Site VPN? para obtener más información. Si lo desea, puede acelerar la conexión de Site-to-Site VPN. Consulte Conexiones de VPN de sitio a sitio acelerada para obtener más información. S2S VPN también puede utilizar las VIF de tránsito de Direct Connect para evitar que el tráfico atraviese Internet, lo que reduce los costos y permite utilizar direcciones IP privadas. Para obtener más información, consulte VPN de IP privada con AWS Direct Connect.
- Software de VPN de sitio a sitio (VPN administrada por el cliente): con esta opción de conexión de VPN, tiene la responsabilidad de aprovisionar y administrar toda la solución de VPN,

normalmente mediante la ejecución de software de VPN en una instancia de EC2. Para obtener más información, consulte VPN de sitio a sitio por software.

Ambas opciones requieren compatibilidad en el dispositivo de puerta de enlace de cliente para terminar el extremo en las instalaciones de los túneles de VPN. Este dispositivo puede ser físico o de software. Para obtener más información sobre los dispositivos de red probados por AWS, consulte la lista de dispositivos de puerta de enlace de cliente probados.

Transit Gateway Connect (TGW Connect)

Transit Gateway Connect utiliza túneles GRE entre AWS Transit Gateway y un dispositivo de puerta de enlace en las instalaciones. BGP se utiliza sobre TGW Connect para permitir el enrutamiento dinámico. Tenga en cuenta que TGW Connect no está cifrado. Para obtener más información, consulte Transit Gateway Connect.

Servicios de conectividad híbrida de AWS

Los servicios de conectividad híbrida de AWS proporcionan componentes de red de alta escalabilidad y disponibilidad. Desempeñan un papel esencial en la creación de soluciones de redes híbridas. En este momento, existen tres puntos de conexión de servicios principales:

- AWS Virtual Private Gateway (VGW) es un servicio regional con una elevada redundancia que proporciona enrutamiento y reenvío de IP en el nivel de VPC, actuando como puerta de enlace para que la VPC se comunique con sus dispositivos de puerta de enlace de cliente. VGW puede terminar las conexiones de AWS S2S VPN y las VIF privadas de AWS Direct Connect.
- AWS Transit Gateway (TGW) es un servicio regional, altamente disponible y escalable que le permite conectar varias VPC entre sí, así como sus redes en las instalaciones a través de VPN de sitio a sitio o Direct Connect mediante una sola puerta de enlace centralizada. Conceptualmente, AWS Transit Gateway actúa como un enrutador de nube virtual redundante y de alta disponibilidad. AWS Transit Gateway admite el enrutamiento de varias rutas de igual costo (ECMP) a través de varias conexiones de Direct Connect, túneles VPN o pares de TGW Connect. Las puertas de enlace de tránsito pueden comunicarse entre sí, tanto en la misma región como entre regiones, lo que permite que sus recursos conectados se comuniquen a través de los enlaces de emparejamiento. Para obtener más información, consulte los escenarios de AWS Transit Gateway.
- Nube de AWS WAN proporciona un panel central para establecer conexiones entre sus sucursales, centros de datos y VPC de Amazon, con lo que se crea una red global con solo unos

clics. Utilice políticas de red para automatizar la administración de la red y las tareas de seguridad en un mismo lugar. Para obtener más información, consulte la documentación de Nube de AWS WAN.

- Direct Connect Gateway (DXGW) es un servicio disponible en todo el mundo que distribuye información de enrutamiento a través de sus conexiones. Tiene un comportamiento similar a los reflectores de ruta BGP en una red tradicional. Los datos no pasan a través de una DXGW, solo gestiona la información de enrutamiento. Puede crear una DXGW en cualquier Región de AWS y acceder a ella desde todas las demás Regiones de AWS. Puede conectar las VIF de Direct Connect a una DXGW y, a continuación, asociar la DXGW a una VGW (con VIF privadas) o a una AWS Transit Gateway (con VIF de tránsito). Consulte Puertas de enlace de Direct Connect para obtener más información. No es necesario crear varias DXGW para la redundancia, ya que se trata de un servicio de disponibilidad global. No obstante, puede elegir utilizar varias DXGW para separar dominios de enrutamiento, por ejemplo, una red de producción y otra de pruebas que desee mantener completamente aisladas.

Consideraciones sobre el tipo de conectividad híbrida y su diseño

En esta sección del documento técnico se tratan las consideraciones que afectan a sus elecciones a la hora de seleccionar una red híbrida para conectar sus entornos en las instalaciones a AWS. Sigue un proceso de pensamiento lógico para ayudarlo a seleccionar una solución óptima de conectividad híbrida. Las consideraciones que afectan su diseño se clasifican en consideraciones que afectan su tipo de conectividad y consideraciones que afectan su diseño de conectividad. Las consideraciones sobre el tipo de conectividad lo ayudarán a decidir entre utilizar una VPN basada en Internet o Direct Connect. Las consideraciones sobre el diseño de la conectividad lo ayudarán a decidir cómo establecer las conexiones.

Se tratan las siguientes consideraciones que afectan su tipo de conectividad: tiempo de implementación, seguridad, SLA, rendimiento y costo. Después de revisar esas consideraciones, y cómo afectan sus opciones de diseño, podrá decidir si utilizar una conexión basada en Internet o Direct Connect es lo más recomendable para satisfacer sus necesidades.

Se tratan las siguientes consideraciones que afectan su diseño de conectividad: escalabilidad, modelo de comunicación, fiabilidad e integración de SD-WAN de terceros. Después de revisar esas consideraciones, y cómo afectan sus opciones de diseño, podrá decidir el diseño lógico óptimo recomendado para satisfacer sus necesidades.

La siguiente estructura se utiliza para debatir y analizar cada una de las consideraciones relativas a la selección y el diseño:

- Definición: breve definición de lo que es la consideración.
- Preguntas clave: incluye un conjunto de preguntas que le permitirán recopilar los requisitos asociados a la consideración.
- Capacidades que se deben tener en cuenta: soluciones para abordar los requisitos asociados a la consideración.
- Árbol de decisiones: para algunas consideraciones o un grupo de consideraciones, se proporciona un árbol de decisiones para ayudarlo a seleccionar la solución de red híbrida óptima.

Las consideraciones que afectan el diseño de su red híbrida se tratan en un orden en el que el resultado de una consideración forma parte de la entrada de la consideración posterior. Como se

ilustra en la figura 2, el primer paso es decidir el tipo de conectividad, seguido de su ajuste con las consideraciones de selección de diseño.

En la figura 2 se muestran las dos categorías de consideraciones, las consideraciones individuales y el orden lógico en el que se tratan las consideraciones en las subsecciones posteriores. Estas son las consideraciones esenciales a la hora de tomar una decisión sobre un diseño de red híbrida. Si el diseño previsto no requiere todas estas consideraciones, puede centrarse en las que se apliquen a sus requisitos.

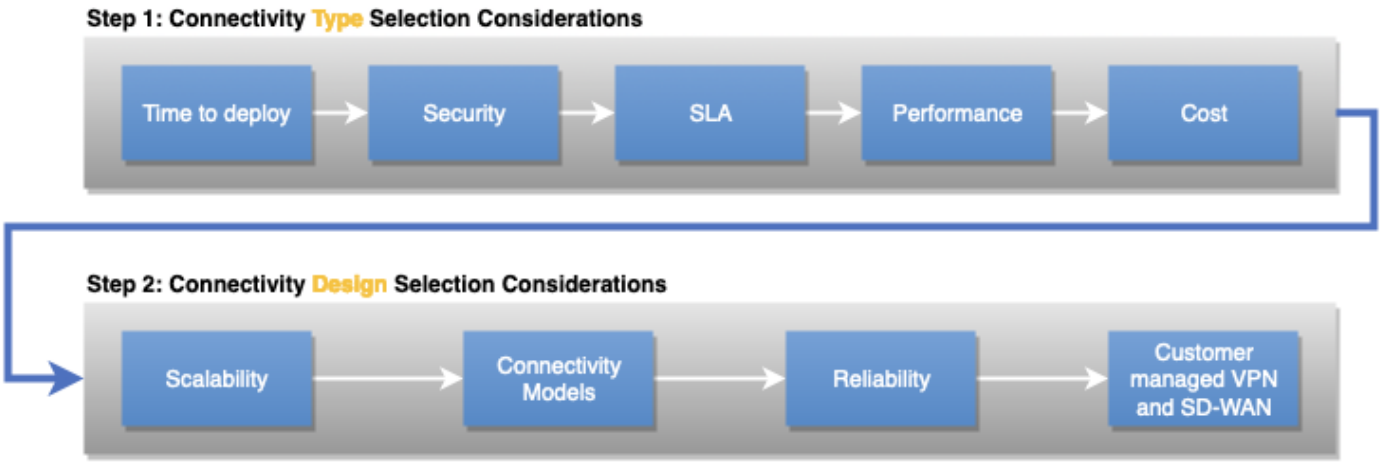

Figura 2. Categorías de consideraciones, consideraciones individuales y orden lógico entre ellas

Selección del tipo de conectividad

En esta sección se describen las consideraciones que afectan el tipo de conectividad que seleccione para su carga de trabajo. Se incluyen el tiempo de implementación, la seguridad, el SLA, el rendimiento y el costo.

Consideraciones

- Momento de implementación
- Seguridad
- Acuerdo de nivel de servicio (SLA)
- Desempeño
- Costo

Momento de implementación

Definición

El momento de implementación puede ser un factor importante a la hora de seleccionar un tipo de conectividad adecuado para una carga de trabajo. En función del tipo de conectividad y de las ubicaciones en las instalaciones, la conectividad puede establecerse en cuestión de horas; no obstante, pueden pasar semanas o meses si hay que instalar circuitos adicionales. Esto influirá en su decisión de utilizar una conexión basada en Internet, una conexión dedicada privada o una conexión alojada privada proporcionada como servicio administrado por un socio de AWS Direct Connect.

Preguntas clave

- ¿Cuál es el plazo necesario para la implementación: horas, días, semanas o meses?
- ¿Cuánto tiempo se necesitará la conexión: será un proyecto de corta duración o una infraestructura permanente?

Capacidades que se deben tener en cuenta

Cuando necesite conectividad de AWS en cuestión de horas o días, lo más probable es que tenga que utilizar una conexión de red existente. Esto significa a menudo establecer una conexión de VPN a AWS a través del Internet público. Si un socio de AWS DX existente le está proporcionando conectividad de AWS privada, podría aprovisionarse una nueva conexión hospedada en cuestión de horas.

Cuando disponga de días o semanas, puede trabajar con un socio de AWS Direct Connect para establecer una conectividad privada a AWS. AWS Direct Connect Los socios lo ayudan a establecer la conectividad de red entre las ubicaciones de AWS Direct Connect y su centro de datos, oficina o entorno de coubicación. Algunos socios de AWS Direct Connect están autorizados a ofrecer conexiones alojadas de Direct Connect. A menudo, las conexiones alojadas pueden aprovisionarse más rápidamente que las dedicadas. AWS Direct Connect El socio aprovisionará cada conexión alojada mediante su infraestructura existente conectada a la red troncal de AWS.

Cuando disponga de varias semanas o meses, puede investigar la posibilidad de establecer una conexión privada dedicada con AWS. Los proveedores de servicios y los socios de AWS Direct Connect facilitan las conexiones dedicadas de AWS Direct Connect. Es habitual que los proveedores de servicios instalen equipos de red en las instalaciones del cliente para facilitar una conexión

dedicada de Direct Connect. Dependiendo del proveedor de servicios, la ubicación de su sitio y otros factores físicos, la instalación de una conexión dedicada de Direct Connect puede llevar desde varias semanas hasta algunos meses.

Si ya tiene sus equipos de red instalados en la misma instalación de coubicación en la que se encuentra la ubicación de AWS Direct Connect, puede establecer rápidamente una conexión dedicada de AWS Direct Connect a través de una conexión cruzada en el sitio de coubicación. Una vez que haya solicitado la conexión, AWS pone a su disposición una Carta de autorización y Asignación de instalaciones de conexión (LOA-CFA) que puede descargar, o le enviará un correo electrónico solicitándole más información. El documento LOA-CFA es la autorización para conectarse a AWS y el proveedor de coubicación o el proveedor de red la necesitan para solicitar una conexión de red por usted.

Tabla 1: Comparación de rentabilidad

	Conectividad basada en Internet	Conexión dedicada de DX (equipo existente en la ubicación de DX)	Conexión dedicada de DX (completamente nueva)	Conexión alojada de DX (puerto existente con socio de DX)	Conexión alojada de DX (completamente nueva)
Tiempo de aprovisionamiento	De horas a días	Days	De varias semanas a meses	De horas a días	Varios días, semanas o meses

> **Note**
>
> Las directrices de tiempo de aprovisionamiento proporcionadas se basan en la observación del mundo real y solo sirven como ilustración. Si se tiene en cuenta la ubicación de su sitio, la proximidad a ubicaciones de Direct Connect y la infraestructura preexistente, todo ello repercutirá en el tiempo de aprovisionamiento. Su socio de AWS Direct Connect lo aconsejará sobre el momento preciso de aprovisionamiento.

Seguridad

Definición

Los requisitos de seguridad influirán en su tipo de conectividad híbrida. Estas consideraciones incluyen:

- Tipo de transporte: conexión a Internet o a una red privada
- Requisitos de cifrado

Preguntas clave

- ¿Permiten sus requisitos y políticas de seguridad utilizar conexiones cifradas a través de Internet para conectarse a AWS u obligan a utilizar conexiones de red privada?
- Cuando se aprovechan las conexiones de red privada, ¿la capa de red tiene que proporcionar cifrado en tránsito?

Soluciones técnicas

Es posible que sus requisitos y políticas de seguridad permitan utilizar Internet o exijan el uso de una conexión de red privada entre AWS y la red de su empresa. También afectan la decisión de si la red debe proporcionar cifrado en tránsito o si es aceptable realizar el cifrado en la capa de aplicación.

Si puede aprovechar Internet, AWS Site-to-Site VPN puede utilizarse para crear túneles cifrados entre su red y sus Amazon VPC o AWS Transit Gateway a través de Internet. La ampliación de su solución SD-WAN a AWS mediante Internet también es una opción si está aprovechando una conexión basada en Internet. La sección VPN y SD-WAN administradas por el cliente, más adelante en este documento técnico, trata las consideraciones específicas para SD-WAN.

Si necesita una conexión de red privada entre AWS y la red de su empresa, AWS recomienda utilizar conexiones dedicadas o conexiones alojadas de AWS Direct Connect. Si se requiere el cifrado en tránsito a través de una conexión de red privada, deberá establecer una VPN a través de Direct Connect (a través de las VIF públicas o las VIF de tránsito) o considerar la posibilidad de utilizar MACsec en una conexión dedicada de 10 Gbps o 100 Gbps.

Tabla 2. Requisitos de tipo de conectividad de Example Automotive Corp

	Site-to-Site VPN	Direct Connect
Transporte	Internet	Conexión de red privada
Cifrado en tránsito	Sí	Requiere S2S VPN sobre DX, S2S VPN sobre una VIF de tránsito o MACsec en una conexión dedicada de 10 Gbps o 100 Gbps.

Acuerdo de nivel de servicio (SLA)

Definición

Las organizaciones empresariales suelen exigir a un proveedor de servicios que cumpla un SLA por cada servicio que la organización consume. A su vez, la organización basa sus servicios en ellos y puede ofrecer a sus consumidores un acuerdo de nivel de servicio. El SLA es importante, ya que describe cómo se presta y funciona el servicio, y a menudo incluye características específicas medibles, como la disponibilidad. Si el servicio incumple el SLA definido, el proveedor de servicios suele ofrecer una compensación económica especificada en el acuerdo. Un SLA define el tipo de medida, el requisito y el periodo de medición. Como ejemplo, consulte la definición del objetivo de tiempo de actividad en el SLA de AWS Direct Connect.

Preguntas clave

- ¿Se requiere un SLA de conexión de conectividad híbrida con créditos de servicio?
- ¿Toda la red híbrida debe cumplir un objetivo de tiempo de actividad?

Capacidades que se deben tener en cuenta

Tipo de conectividad: la conectividad a Internet puede ser impredecible. Aunque AWS tiene mucho cuidado con los múltiples enlaces establecidos con un conjunto diverso de ISP, la administración de Internet está simplemente fuera de AWS o del dominio administrativo de un único proveedor. Existe una cantidad limitada de ingeniería de rutas e influencia en el tráfico que un proveedor de nube puede hacer una vez que el tráfico ha salido de la frontera de su red. Dicho esto, existe un SLA de AWS Site-to-Site VPN que proporciona objetivos de disponibilidad para los puntos de conexión de AWS Site-to-Site VPN.

AWS Direct Connect ofrece un SLA formal con créditos de servicio calculados como un porcentaje del total de los cargos por hora de puerto de AWS Direct Connect que ha pagado por las conexiones aplicables que no estén disponibles durante el ciclo de facturación mensual en el que no se haya cumplido el SLA. Este es el transporte recomendado si se requiere un SLA. AWS Direct Connect enumera los requisitos mínimos de configuración específicos para cada objetivo de tiempo de actividad, como el número de ubicaciones, conexiones y otros detalles de configuración de AWS Direct Connect. El incumplimiento de los requisitos significa que no se pueden ofrecer créditos de servicio en caso de que el servicio incumpla los SLA definidos.

Es importante destacar que, aunque el servicio seleccionado para proporcionar conectividad híbrida esté configurado para cumplir los requisitos del SLA, es posible que el resto de la red no proporcione el mismo nivel de SLA. La responsabilidad de AWS termina en la ubicación de AWS Direct Connect en el puerto de AWS Direct Connect. Una vez que AWS transfiere el tráfico a la red de su organización, deja de ser responsabilidad de AWS. Si utiliza un proveedor de servicios entre AWS y su red en las instalaciones, la conectividad está sujeta al SLA entre usted y el proveedor de servicios, si procede. Tenga en cuenta que toda la red híbrida es tan buena como su parte más débil a la hora de diseñar la conectividad híbrida.

Los socios de AWS Direct Connect ofrecen conectividad de AWS Direct Connect. El socio puede ofrecer un SLA con créditos de servicio basado en su oferta de productos hasta el punto de demarcación con AWS. Esta opción debe evaluarse e investigarse más a fondo directamente con los socios de APN. AWS publica una lista de socios de entrega validados.

Diseño lógico: además del tipo de conectividad, también debe tener en cuenta otros componentes como parte de su diseño general. A modo de ejemplo, AWS Transit Gateway tiene su propio SLA, al igual que AWS S2S VPN. Puede que utilice AWS Transit Gateway para escalar y AWS S2S VPN por motivos de seguridad, pero debe diseñar ambos de forma coherente con cada SLA a fin de reunir los requisitos para obtener créditos de servicio con cada servicio respectivo.

Consulte Recomendaciones de resiliencia de AWS Direct Connect y Kit de herramientas de resiliencia.

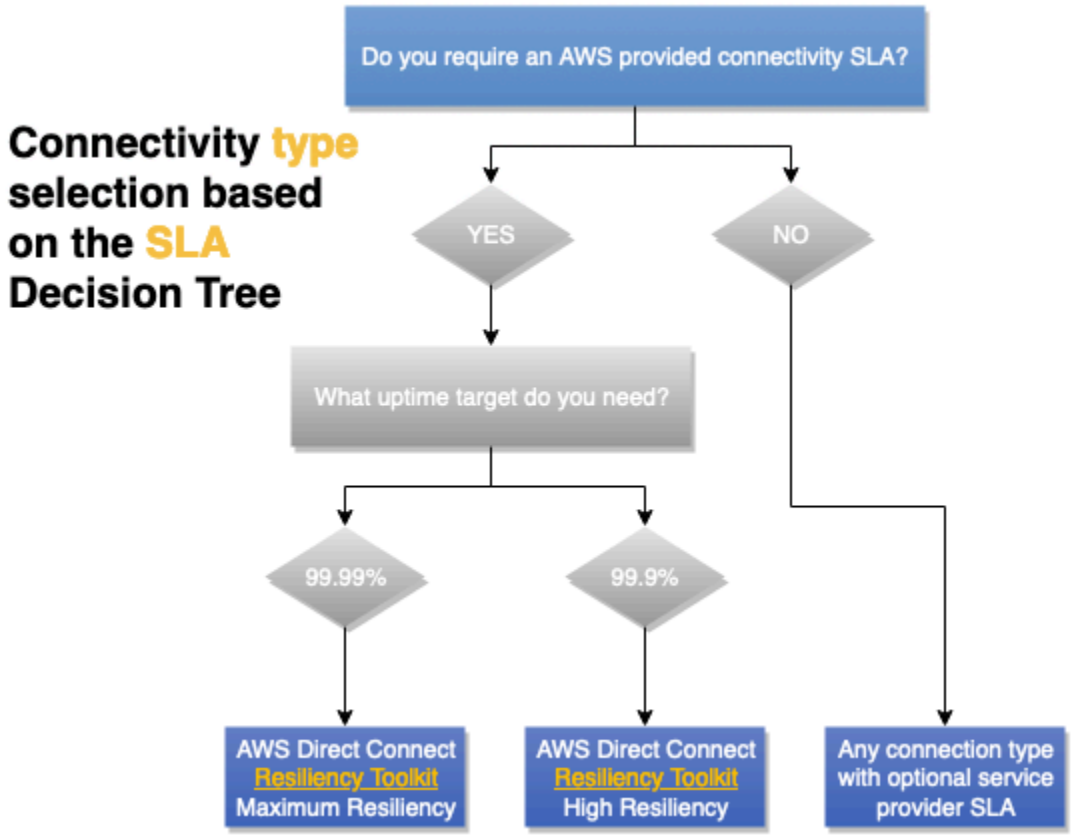

Figura 3. Árbol de decisiones para la consideración del SLA

Desempeño

Definición

Hay varios factores que influyen en el rendimiento de la red, como la latencia, la pérdida de paquetes, la fluctuación y el ancho de banda. La importancia de cada uno de estos factores puede variar en función de los requisitos de la aplicación.

Preguntas clave

Basándose en los requisitos de su aplicación, necesita identificar y priorizar los factores de rendimiento de la red que afectan el comportamiento de su aplicación y a la experiencia del usuario.

Ancho de banda

El ancho de banda se refiere a la velocidad de transferencia de datos de una conexión y suele medirse en bits por segundo (bps). Los megabits por segundo (Mbps) y los gigabits por segundo

(Gbps) son escalas comunes, y son de base 10 (1 000 000 de bits por segundo = 1 Mbps) en contraposición a base 2 (2^10) vista en otras partes.

Cuando evalúe las necesidades de ancho de banda de las aplicaciones, tenga en cuenta que estas pueden cambiar con el tiempo. La implementación inicial en la nube, las operaciones normales, las nuevas cargas de trabajo y los escenarios de conmutación por error pueden tener diferentes requisitos de ancho de banda.

Las aplicaciones pueden tener sus propias consideraciones de ancho de banda. Algunas aplicaciones pueden requerir un rendimiento determinista sobre una conexión de ancho de banda alto, mientras que otras pueden requerir tanto un rendimiento determinista como un ancho de banda alto. Una aplicación puede necesitar una configuración especial para utilizar múltiples flujos de tráfico (a veces denominados flujos o sockets) en paralelo si está alcanzando los límites de ancho de banda por flujo de tráfico, lo que le permite utilizar más ancho de banda de la conexión. Las VPN pueden limitar el rendimiento debido a la sobrecarga de los túneles, a los límites inferiores de las MTU o a las limitaciones de ancho de banda de hardware.

Latency (Latencia)

La latencia es el tiempo necesario para que un paquete vaya del origen al destino a través de una conexión de red, y suele medirse en milisegundos (ms), con requisitos de latencia bajos que a veces se expresan en microsegundos (µs). La latencia es una función de la velocidad de la luz, por lo que la latencia aumenta con la distancia.

Los requisitos de latencia de las aplicaciones pueden adoptar diferentes formas. Una aplicación muy interactiva, como un escritorio virtual, puede tener un objetivo de latencia medido desde que el usuario realiza una entrada hasta que ve que el escritorio virtual reacciona a esa entrada. Las aplicaciones de voz sobre IP (VoIP) pueden tener requisitos similares. Un segundo tipo de carga de trabajo que se debe tener en cuenta son las que son altamente transaccionales, que necesitan una respuesta del servidor antes de poder continuar. Las bases de datos u otras formas de almacenes de claves/valores pueden verse muy afectadas por el aumento de la latencia de red.

Jitter (Irregularidad)

La fluctuación mide la coherencia de la latencia de la red y, al igual que la latencia, suele medirse en milisegundos (ms).

Los requisitos de fluctuación de la aplicación se encuentran normalmente en aplicaciones de streaming en tiempo real, como la entrega de vídeo y voz. Estas aplicaciones suelen requerir que

su flujo de datos tenga una velocidad y un retardo coherentes, con pequeños búferes para corregir pequeños volúmenes de fluctuación.

Pérdida de paquetes

La pérdida de paquetes es la medida del porcentaje de tráfico de red que no se entrega. Todas las redes tienen en ocasiones cierto grado de pérdida de paquetes debido a las grandes ampliaciones de tráfico, las reducciones de capacidad, los errores en los equipos de red y otras razones. Por lo tanto, las aplicaciones deben tener cierta tolerancia a la pérdida de paquetes. No obstante, su tolerancia puede variar de una aplicación a otra.

Las aplicaciones que utilizan TCP para transportar su tráfico tienen la capacidad de corregir la pérdida de paquetes mediante la retransmisión. Las aplicaciones que utilizan UDP o sus propios protocolos sobre IP necesitan implementar sus propios medios para gestionar la pérdida de paquetes y pueden ser muy sensibles a ella. Una aplicación de voz sobre IP puede simplemente insertar silencio en la parte de la llamada que tuvo la pérdida de paquetes, en lugar de intentar una retransmisión. Algunas soluciones de VPN incluyen sus propios mecanismos para recuperarse de la pérdida de paquetes en la red que utilizan para transportar el tráfico.

Capacidades que se deben tener en cuenta

Cuando se requiere una latencia y un rendimiento predecibles, AWS Direct Connect es la opción recomendada, ya que proporciona un rendimiento determinista. El ancho de banda puede seleccionarse en función de los requisitos de rendimiento. AWS recomienda utilizar AWS Direct Connect cuando necesite una experiencia de red más coherente que la que pueden proporcionar las conexiones basadas en Internet. Las VIF privadas y las VIF de tránsito admiten tramas gigantes, que pueden reducir el número de paquetes que recorren la red y pueden mejorar el rendimiento gracias a la reducción de la sobrecarga. AWS Direct Connect SiteLink permite utilizar la red troncal de AWS para proporcionar conectividad entre sus ubicaciones y puede habilitarse bajo demanda. El ancho de banda utilizado para SiteLink debe tenerse en cuenta para la selección del ancho de banda de Direct Connect.

Utilizar una VPN sobre AWS Direct Connect agrega cifrado. No obstante, reduce el tamaño de la MTU, lo que podría reducir el rendimiento. En la documentación de AWS Site-to-Site VPN encontrará las capacidades de Site-to-Site (S2S) VPN administrada por AWS. Muchas ubicaciones de Direct Connect admiten MACsec si el cifrado a través de su conexión es el principal requisito de cifrado. MACsec no tiene las mismas consideraciones de MTU o de rendimiento potencial que las conexiones de VPN de sitio a sitio. AWS Transit Gateway permite a los clientes escalar

horizontalmente el número de conexiones de VPN y aumentar el rendimiento en consecuencia con el enrutamiento multitrayectoria de costo equivalente (ECMP). VPN de sitio a sitio administrada por AWS admite la utilización de las VIF de tránsito de Direct Connect para la conectividad privada. Consulte VPN de IP privada con AWS Direct Connect para obtener detalles.

Otra opción es utilizar una VPN de sitio a sitio administrada por AWS a través de Internet. Puede ser una opción atractiva debido a su bajo costo y su amplia disponibilidad. No obstante, tenga en cuenta que el rendimiento a través de Internet es la mejor acción. Los fenómenos meteorológicos de Internet, la congestión y el aumento de los periodos de latencia pueden ser impredecibles. AWS ofrece una solución con AWS S2S VPN acelerada, que puede mitigar algunos de los inconvenientes de utilizar una ruta de Internet. S2S VPN acelerada utiliza AWS Global Accelerator, que permite que el tráfico de VPN entre en la red de AWS lo antes y lo más cerca posible del dispositivo de puerta de enlace de cliente. De este modo se optimiza la ruta de la red, mediante la red global de AWS libre de congestiones, para dirigir el tráfico al punto de conexión que proporcione el mejor rendimiento. Puede utilizar conexiones de VPN acelerada para evitar las interrupciones en la red que podrían producirse cuando el tráfico se direcciona a través del Internet público.

Costo

Definición

En la nube, el costo de la conectividad híbrida incluye el costo de los recursos aprovisionados y el uso. El costo de los recursos aprovisionados se mide en unidades de tiempo, normalmente por hora. El uso se refiere a la transferencia y el procesamiento de datos, que suele medirse en gigabytes (GB). Otros costos incluyen el costo de la conectividad al punto de presencia de red de AWS. Si su red se encuentra en la misma instalación de coubicación, podría ser tan poco como el costo de una conexión cruzada. Si su red se encuentra en un lugar diferente, habrá que contar con los costos de un proveedor de servicios o de un socio de APN Direct Connect.

Preguntas clave

- ¿Cuántos datos prevé enviar a AWS al mes desde sus instalaciones y desde Internet?
- ¿Cuántos datos prevé enviar desde AWS al mes a sus instalaciones y a Internet?
- ¿Con qué frecuencia cambiarán estas cantidades?
- ¿Qué cambia en un escenario de error?

Capacidades que se deben tener en cuenta

Si tiene cargas de trabajo con mucho ancho de banda que desea ejecutar en AWS, AWS Direct Connect puede reducir sus costos de red de entrada y salida de AWS de dos maneras. En primer lugar, al transferir los datos de ida y vuelta de AWS directamente, puede reducir los costos de ancho de banda que paga a su proveedor de servicios de Internet. En segundo lugar, todos los datos transferidos a través de su conexión dedicada se cobran con la tarifa de transferencia de datos de AWS Direct Connect reducida, en lugar de con las tarifas de transferencia de datos de Internet. Consulte la página de precios de Direct Connect para obtener detalles.

AWS Direct Connect permite utilizar AWS Direct Connect SiteLink para interconectar sus sitios mediante la red troncal de AWS. Consulte el blog de lanzamiento de SiteLink para obtener más información. El uso de esta capacidad conlleva los costos normales de transferencia de datos de Direct Connect, junto con un cargo por hora de activación de SiteLink. Puede activar y desactivar SiteLink bajo demanda y puede ser una buena opción para escenarios de error que impliquen la conectividad a Internet o a una red privada.

Si utiliza un proveedor de servicios de red para la conectividad entre las instalaciones y una ubicación de Direct Connect, su capacidad y el tiempo necesario para cambiar sus compromisos de ancho de banda se basan en su contrato con el proveedor de servicios.

La red troncal de AWS puede enviar su tráfico a cualquier Región de AWS, excepto China, desde cualquier punto de presencia de la red de AWS. Esta capacidad tiene muchas ventajas técnicas respecto a utilizar Internet para acceder a Regiones de AWS remotas, pero tiene un costo. Consulte la página de precios de transferencia de datos de EC2 para obtener detalles. Si hay AWS Transit Gateway en la ruta de tráfico, se agrega el costo de procesamiento de datos por GB. No obstante, si utiliza el emparejamiento entre regiones entre dos Transit Gateway, solo se le facturará una vez por el procesamiento de datos de Transit Gateway.

El diseño óptimo de la aplicación mantiene el procesamiento de datos en AWS y minimiza los gastos innecesarios de salida de datos. La entrada de datos a AWS es gratis.

> **Note**
>
> Como parte de la solución global de conectividad, además del costo de la conexión de AWS, también debe considerar el costo de la conectividad de extremo a extremo, incluido el costo del proveedor de servicios, las conexiones cruzadas, los bastidores y el equipo en la ubicación de DX (si es necesario).

Si no está seguro de si debe utilizar Internet o una conexión privada, calcule un punto de equilibrio en el que AWS Direct Connect resulte menos costoso que utilizar Internet. Si el volumen de datos significa que AWS Direct Connect es menos costoso, y necesita una conectividad permanente, AWS Direct Connect es la opción de conectividad óptima.

Si la conectividad es temporal e Internet cumple otros requisitos, puede resultar más barato utilizar AWS S2S VPN a través de Internet debido a la elasticidad de Internet. Tenga en cuenta que esto requiere que disponga de suficiente conectividad a Internet desde su red en las instalaciones.

Si se encuentra en una instalación que disponga de AWS Direct Connect (la lista está disponible en el sitio web de Direct Connect), puede establecer una conexión cruzada con AWS. Esto significa utilizar conexiones dedicadas a 1, 10 o 100 Gbps. Los socios de AWS Direct Connect ofrecen más opciones de ancho de banda y capacidades menores, lo que puede optimizar su costo de conectividad. Por ejemplo, puede empezar con una conexión alojada de 50 Mbps en lugar de una conexión dedicada de 1 Gbps.

Con AWS Transit Gateway, puede compartir sus conexiones de VPN y Direct Connect con muchas VPC. Aunque se le cobrará por el número de conexiones que realice a AWS Transit Gateway por hora y por la cantidad de tráfico que fluya a través de AWS Transit Gateway, simplifica la administración y reduce el número de conexiones de VPN y las VIF necesarias. Los beneficios y el ahorro de costos de una menor carga operativa pueden compensar fácilmente el costo adicional del procesamiento de datos. Opcionalmente, puede considerar un diseño en el que se encuentre AWS Transit Gateway en la ruta de tráfico para la mayoría de las VPC, pero no para todas. Este enfoque evita las tasas de procesamiento de datos de AWS Transit Gateway para los casos de uso en los que es necesario transferir grandes cantidades de datos a AWS. Consulte la sección Modelos de conectividad para obtener más detalles sobre este diseño. Otro enfoque consiste en combinar AWS Direct Connect como ruta principal con AWS S2S VPN a través de Internet como ruta de copia de seguridad o conmutación por error. Aunque es técnicamente factible y muy rentable, esta solución tiene inconvenientes técnicos (analizados en la sección Fiabilidad de este documento técnico) y puede ser más difícil de administrar. AWS no la recomienda para cargas de trabajo muy críticas o críticas.

El enfoque final es una VPN o SD-WAN administrada por el cliente e implementada en las instancias de Amazon EC2. Esto puede ser más barato a escala si hay de decenas a cientos de sitios en comparación con AWS S2S VPN. No obstante, hay que tener en cuenta la sobrecarga de administración, los costos de las licencias y el costo de los recursos de EC2 para cada dispositivo virtual.

Matriz de decisiones

Table 3. Entradas de diseño de conectividad de Example Corp. Automotive

Categoría	VPN o SD-WAN administrada por el cliente	AWS S2S VPN	AWS S2S VPN acelerada	Conexión alojada de AWS Direct Connect	Conexión dedicada de AWS Direct Connect
Requiere conexión a Internet	Sí	Sí	Sí	No	No
Costo de los recursos aprovisionados	Licencias de instancia de EC2 y de software	AWS S2S VPN	AWS S2S VPN y AWS Global Accelerator	Sector de capacidad aplicable del costo de puerto	Costo de puerto dedicado
Costo de transferencia de datos	Tarifa de Internet	Tarifa de Internet o tarifa de conexión directa	Internet con transferencia de datos premium	Tarifa de Direct Connect	Tarifa de Direct Connect
Gateway de tránsito	Opcional	Opcional	Obligatorio	Opcional	Opcional
Costo de procesamiento de datos de AWS	N/A	Solo con AWS Transit Gateway	Sí	Solo con AWS Transit Gateway	Solo con AWS Transit Gateway
¿Se puede usar sobre AWS Direct Connect?	Sí	Sí	No	N/A	N/A

Selección de diseño de conectividad

En esta sección del documento técnico se tratan las consideraciones que afectan a la selección de su diseño de conectividad. El diseño de conectividad incluye los aspectos lógicos, así como la forma de diseñar y optimizar la fiabilidad de su conectividad híbrida.

Se tratarán las siguientes consideraciones: escalabilidad, modelos de conectividad, fiabilidad y VPN y SD-WAN administradas por el cliente.

Consideraciones

- Escalabilidad
- Modelos de conectividad
- Fiabilidad
- VPN y SD-WAN administradas por el cliente

Escalabilidad

Definición

La escalabilidad hace referencia a la capacidad de su solución de conectividad para crecer y evolucionar con el tiempo a medida que cambian sus requisitos.

A la hora de diseñar una solución, debe tener en cuenta el tamaño actual, así como el crecimiento previsto. Este crecimiento puede ser orgánico o estar relacionado con una rápida expansión, como en el tipo de escenarios de fusión y adquisición.

Nota: en función de la arquitectura de la solución elegida, puede que no sea necesario tener en cuenta todos los elementos anteriores. No obstante, pueden servir como elementos fundamentales para identificar los requisitos de escalabilidad de las soluciones de red híbrida más comunes. Este documento técnico se centra en la selección y el diseño de la conectividad híbrida. Se recomienda tener en cuenta también la escala de la conectividad híbrida con respecto a la arquitectura de red de la VPC. Para obtener más información, consulte el documento técnico Creación de una infraestructura de red de AWS de varias VPC escalable y segura.

Preguntas clave

- ¿Cuál es el número actual y previsto de VPC que requieren conectividad con el sitio o los sitios en las instalaciones?

- ¿Se implementan las VPC en una única Región de AWS o en varias?
- ¿Cuántos sitios en las instalaciones deben conectarse a AWS?
- ¿Cuántos dispositivos de puerta de enlace de cliente (normalmente enrutadores o firewalls) tiene por sitio que necesita conectar a AWS?
- ¿Cuántas rutas se espera que se anuncien a Amazon VPC y cuál es el número de rutas que se espera recibir desde AWS?
- ¿Es necesario aumentar el ancho de banda para AWS con el tiempo?

Capacidades que se deben tener en cuenta

La escala es un factor importante en el diseño de la conectividad híbrida. Para ello, la sección siguiente incorporará la escala como parte del diseño del modelo de conectividad objetivo.

A continuación, se describen las prácticas recomendadas para minimizar la complejidad de escala del diseño de conectividad de red híbrida:

- La integración de rutas debe utilizarse para reducir el número de rutas anunciadas y recibidas de AWS. Por ello, el esquema de direccionamiento IP debe diseñarse para utilizar al máximo el resumen de rutas. La ingeniería de tráfico es una consideración general clave. Para obtener más información sobre la ingeniería de tráfico, consulte la subsección Ingeniería de tráfico de la sección Fiabilidad.
- Minimice el número de sesiones de emparejamiento BGP mediante el uso de DXGW con VGW o AWS Transit Gateway, donde una única sesión BGP puede proporcionar conectividad a varias VPC.
- Considere la WAN en la nube cuando sea necesario conectar entre sí varias Regiones de AWS y sitios en las instalaciones.

Modelos de conectividad

Definición

El modelo de conectividad se refiere al patrón de comunicación entre las redes en las instalaciones y los recursos en la nube de AWS. Puede implementar recursos en la nube en una Amazon VPC en una Región de AWS o varias VPC en múltiples regiones, así como servicios de AWS que tengan un punto de conexión público en una o varias Regiones de AWS, como Amazon S3 y DynamoDB.

Preguntas clave

- ¿Es necesaria la comunicación entre VPC en una región y entre regiones?
- ¿Existe algún requisito para acceder a los puntos de conexión públicos de AWS directamente desde las instalaciones?
- ¿Existe algún requisito para acceder a los servicios de AWS mediante puntos de conexión de VPC desde las instalaciones?

Capacidades que se deben tener en cuenta

A continuación, se presentan algunos de los escenarios más comunes de modelo de conectividad. Cada modelo de conectividad abarca requisitos, atributos y consideraciones.

Nota: Tal y como se destacó anteriormente, este documento técnico se centra en la conectividad híbrida entre las redes en las instalaciones y AWS. Para obtener más detalles sobre el diseño para interconectar las VPC, consulte el documento técnico Creación de una infraestructura de red de AWS de varias VPC escalable y segura.

Modelos

- AWS Site-to-Site VPN acelerada: AWS Transit Gateway, una sola Región de AWS
- AWS DX: DXGW con VGW, una sola región
- AWS DX: DXGW con VGW, varias regiones y emparejamiento público de AWS
- AWS DX: DXGW con AWS Transit Gateway, varias regiones y emparejamiento público de AWS
- AWS DX: DXGW con AWS Transit Gateway, varias regiones (más de tres)

AWS Site-to-Site VPN acelerada: AWS Transit Gateway, una sola Región de AWS

Este modelo se construye a partir de:

- Una sola Región de AWS.
- Conexión de VPN de sitio a sitio administrada por AWS con AWS Transit Gateway.
- VPN acelerada habilitada.

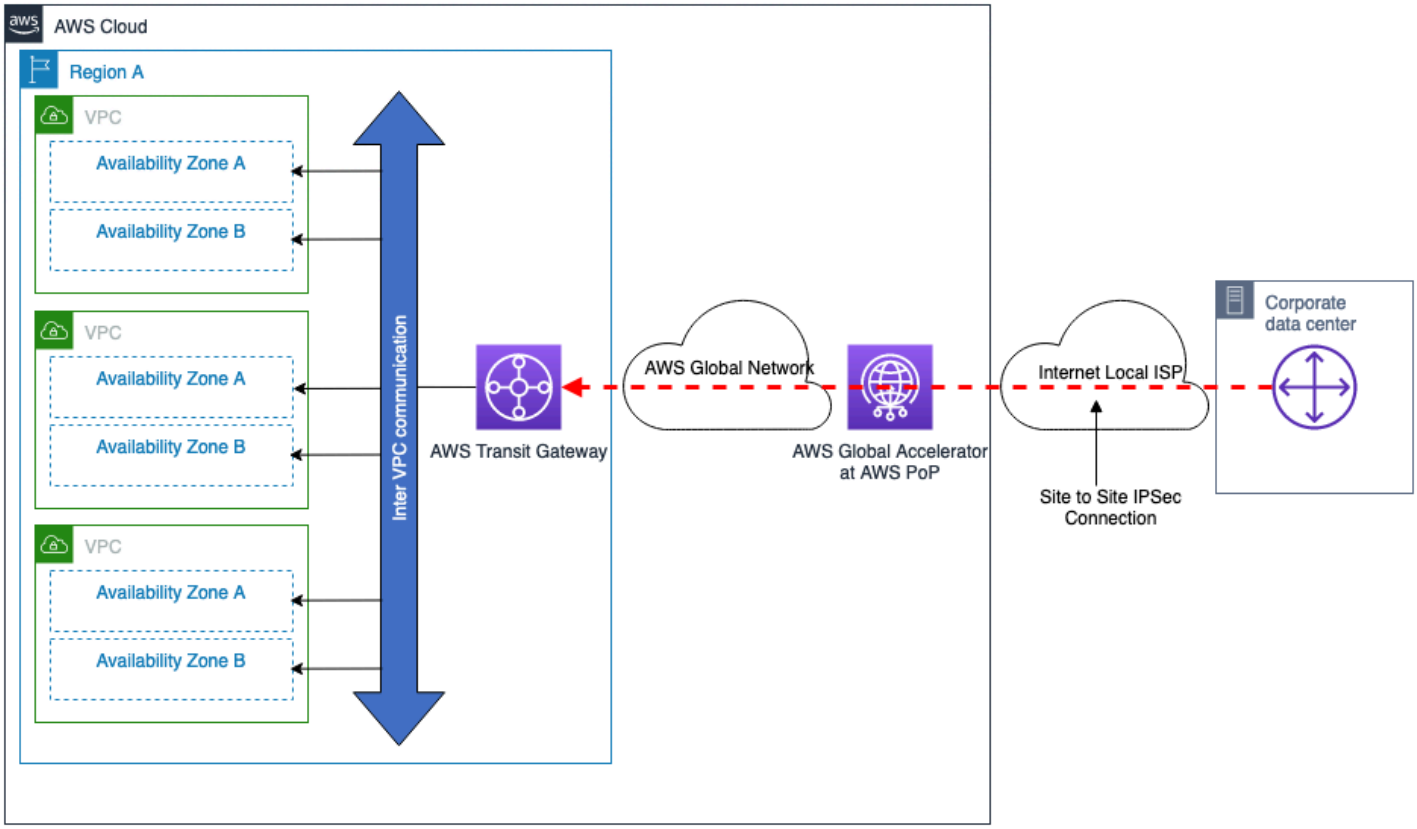

Figura 4. VPN administrada de AWS: AWS Transit Gateway, una sola Región de AWS

Atributos del modelo de conectividad:

- Proporcionar la posibilidad de establecer conexiones de VPN optimizadas a través del Internet público mediante el uso de conexiones de AWS Site-to-Site VPN acelerada.
- Proporcionar la capacidad de lograr un mayor ancho de banda de conexión de VPN mediante la configuración de varios túneles VPN con ECMP.
- Se puede usar para conectarse desde varios sitios remotos.
- Ofrece conmutación por error automatizada con enrutamiento dinámico (BGP).
- Con AWS Transit Gateway conectado a las VPC, todas las VPC conectadas pueden utilizar las mismas conexiones de VPN. También puede controlar el modelo de comunicación que desee entre las VPC. Para obtener más información, consulte Cómo funcionan las puertas de enlace de tránsito.
- Ofrece opciones de diseño flexibles para integrar dispositivos virtuales de seguridad y SD-WAN de terceros con AWS Transit Gateway. Consulte Seguridad de red centralizada para tráfico de VPC a VPC y en las instalaciones a VPC.

Consideraciones de escala:

- Hasta 50 Gbps de ancho de banda con varios túneles IPsec y ECMP configurados (cada flujo de tráfico estará limitado al ancho de banda máximo por túnel VPN).
- Se pueden conectar miles de VPC por AWS Transit Gateway.
- Consulte en Cuotas de VPN de sitio a sitio otros límites de escala, como el número de rutas.

Otras consideraciones:

- Los costos adicionales de procesamiento de AWS Transit Gateway para la transferencia de datos entre el centro de datos en las instalaciones y AWS.
- No se puede hacer referencia a los grupos de seguridad de una VPC remota en AWS Transit Gateway; sin embargo, esto lo admite el emparejamiento de VPC.

AWS DX: DXGW con VGW, una sola región

Este modelo se construye a partir de:

- Una sola Región de AWS.
- Conexiones de AWS Direct Connect dobles a ubicaciones de DX independientes.
- AWS DXGW directamente asociado a las VPC mediante VGW.
- Uso opcional de AWS Transit Gateway para la comunicación entre VPC.

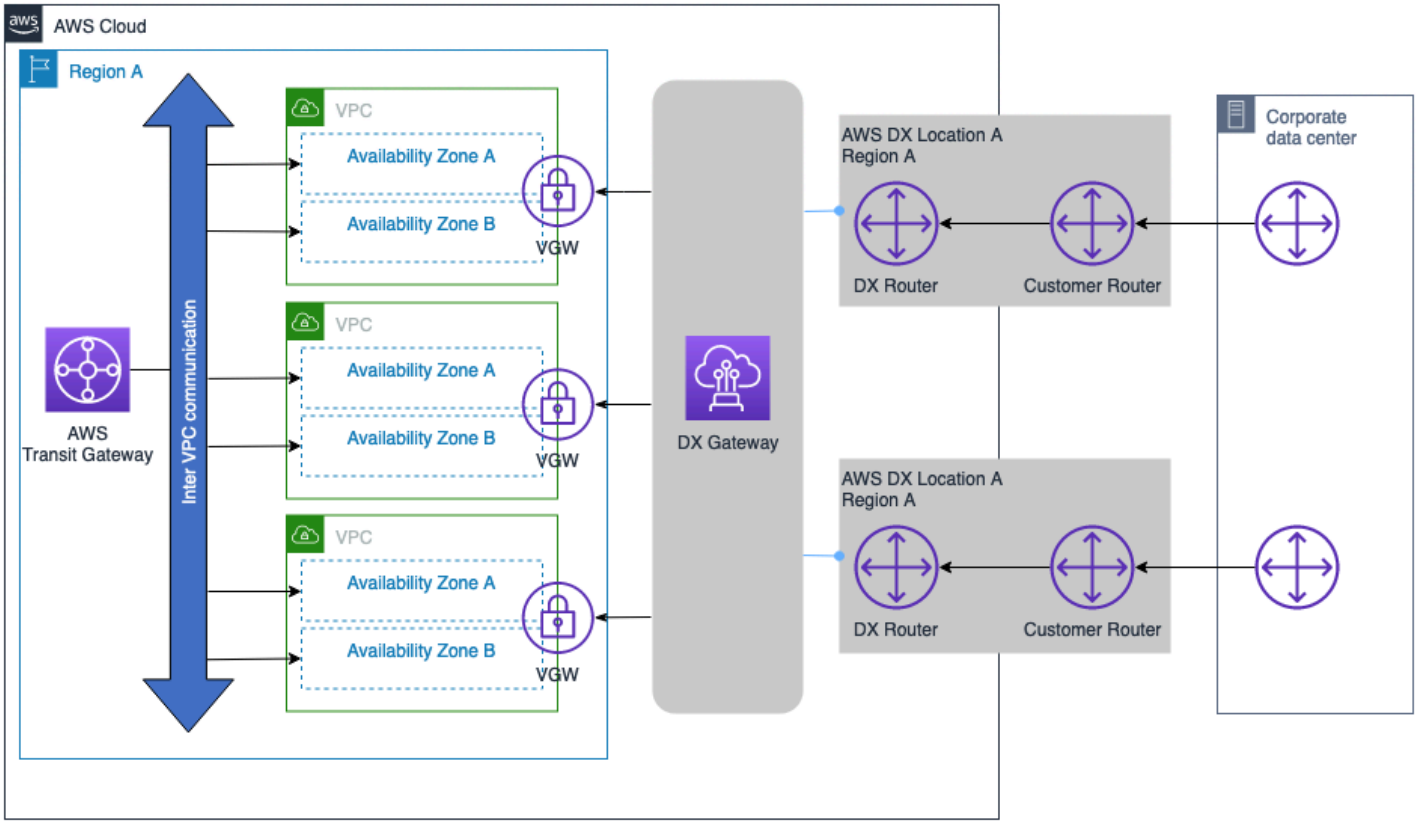

Figura 5. AWS DX: DXGW con VGW, una sola Región de AWS

Atributos del modelo de conectividad:

- Proporciona la capacidad de conectarse a VPC y conexiones de DX en otras regiones en el futuro.
- Ofrece conmutación por error automatizada con enrutamiento dinámico (BGP).
- Con AWS Transit Gateway puede controlar el modelo de comunicación que desea entre las VPC. Para obtener más información, consulte Cómo funcionan las puertas de enlace de tránsito.

Consideraciones de escala:

Consulte las cuotas de AWS Direct Connect para obtener más información sobre otros límites de escalado como, por ejemplo, el número de prefijos admitidos, el número de VIF por tipo de conexión de DX (dedicada o alojada). Algunas consideraciones clave:

- La sesión BGP para una VIF privada puede anunciar hasta 100 rutas cada una para IPv4 e IPv6.

- Se pueden conectar hasta 20 VPC por DXGW en una sola sesión de BGP. Si se necesitan más de 20 VPC, se pueden agregar DXGW adicionales para facilitar la conectividad a escala o considerar la posibilidad de utilizar la integración de Transit Gateway.
- Se pueden agregar AWS Direct Connect adicionales según se desee.

Otras consideraciones:

- No genera costos de procesamiento relacionados con AWS Transit Gateway para la transferencia de datos entre AWS y redes en las instalaciones.
- No se puede hacer referencia a los grupos de seguridad de una VPC remota sobre AWS Transit Gateway (necesita emparejamiento de VPC).
- El emparejamiento de VPC puede utilizarse en lugar de AWS Transit Gateway para facilitar la comunicación entre las VPC. No obstante, esto agrega complejidad operativa para crear y administrar a escala un gran número de emparejamientos de VPC punto a punto.
- Si no es necesaria la comunicación entre VPC, tampoco se necesita AWS Transit Gateway ni el emparejamiento entre VPC en este modelo de conectividad.

AWS DX: DXGW con VGW, varias regiones y emparejamiento público de AWS

Este modelo se construye a partir de:

- Centros de datos múltiples en las instalaciones con conexiones duales a AWS.
- Conexiones de AWS Direct Connect dobles a ubicaciones de DX independientes.
- AWS DXGW asociado directamente a más de 10 VPC mediante VGW, hasta 20 VPC mediante VGW.
- Uso opcional de AWS Transit Gateway para la comunicación entre VPC y entre regiones.

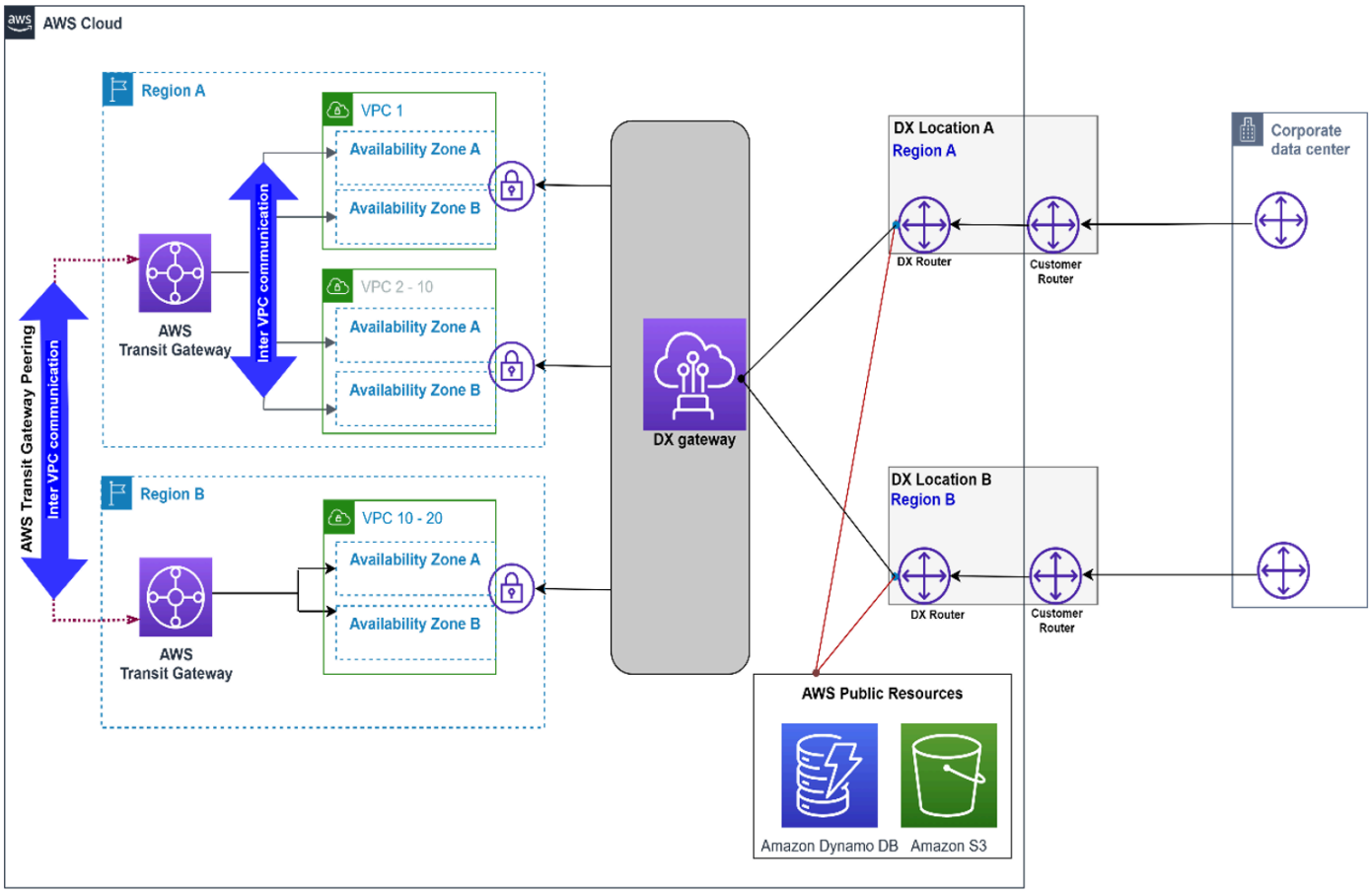

Figura 6. AWS DX: DXGW con VGW, varias regiones y VIF pública

Atributos del modelo de conectividad:

- AWS DXGW asociado directamente a más de 10 VPC mediante VGW, hasta 20 VPC mediante VGW.
- La VIF pública de AWS DX se usa para acceder a servicios públicos de AWS, como Amazon S3, directamente sobre las conexiones de AWS DX.
- Proporcionar la capacidad de conectarse a VPC y conexiones de DX en otras regiones en el futuro.
- Comunicación entre VPC y VPC entre regiones facilitada por el emparejamiento de AWS Transit Gateway y Transit Gateway.

Consideraciones de escala:

Consulte las cuotas de AWS Direct Connect para obtener más información sobre otros límites de escalado como, por ejemplo, el número de prefijos admitidos, el número de VIF por tipo de conexión de DX (dedicada o alojada). Algunas consideraciones clave:

- La sesión BGP para una VIF privada puede anunciar hasta 100 rutas cada una para IPv4 e IPv6.
- Se pueden conectar hasta 20 VPC por DXGW a través de una única sesión BGP en cada VIF privada, hasta 30 VIF privadas por DXGW.
- Se pueden agregar AWS Direct Connect adicionales según se desee.

Otras consideraciones:

- No genera costos de procesamiento relacionados con AWS Transit Gateway para la transferencia de datos entre AWS y redes en las instalaciones.
- No se puede hacer referencia a los grupos de seguridad de una VPC remota mediante AWS Transit Gateway (necesita emparejamiento de VPC).
- El emparejamiento de VPC puede utilizarse en lugar de AWS Transit Gateway para facilitar la comunicación entre las VPC. No obstante, esto agregará complejidad operativa para crear y administrar a escala un gran número de emparejamientos de VPC punto a punto.
- Si no es necesaria la comunicación entre VPC, tampoco se necesita AWS Transit Gateway ni el emparejamiento entre VPC en este modelo de conectividad.

AWS DX: DXGW con AWS Transit Gateway, varias regiones y emparejamiento público de AWS

Este modelo se construye a partir de:

- Varias Regiones de AWS.
- Conexiones de AWS Direct Connect dobles a ubicaciones de DX independientes.
- Centro de datos único en las instalaciones con conexiones duales a AWS.
- AWS DXGW con AWS Transit Gateway.
- Gran escala de VPC por región.

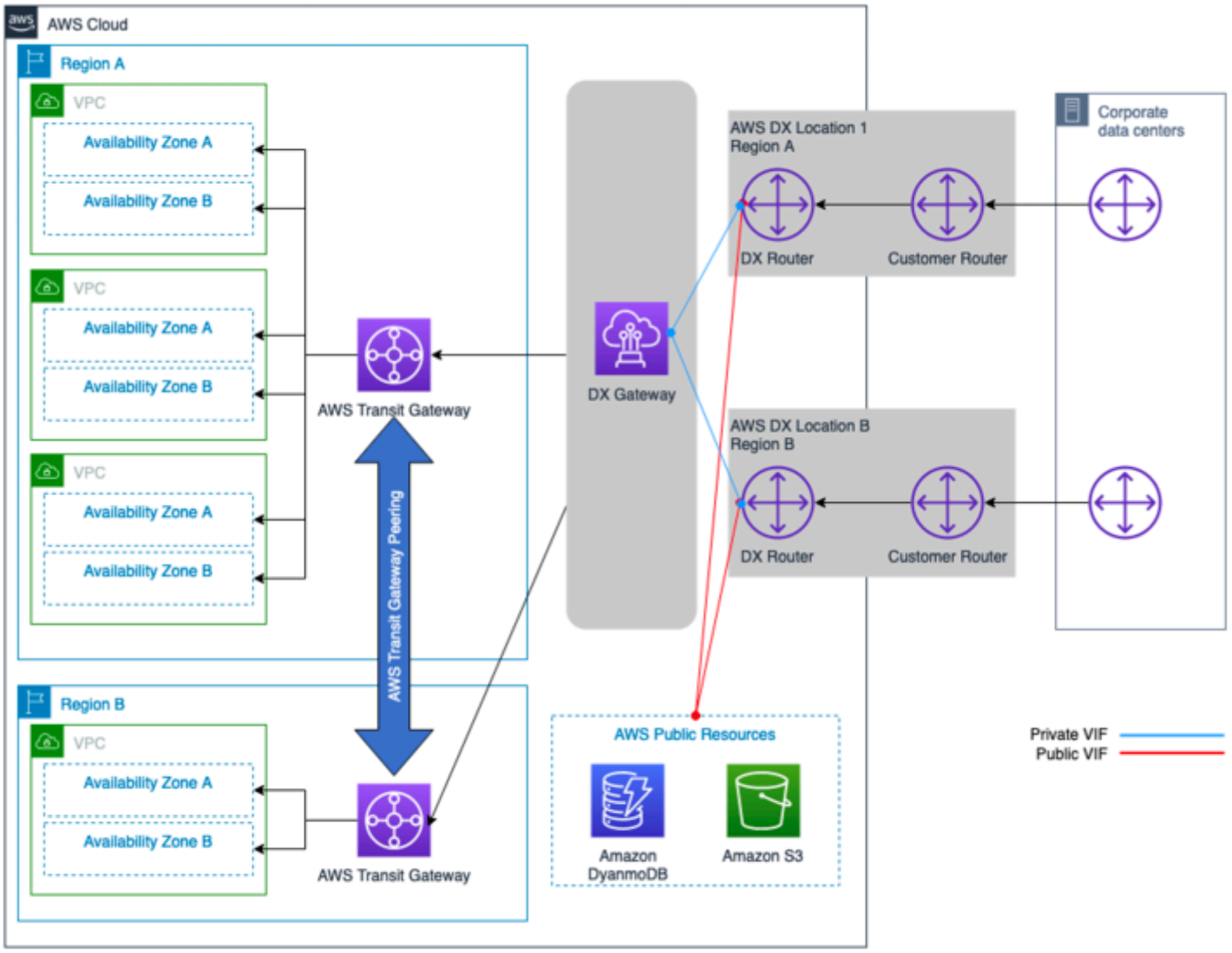

Figura 7. AWS DX: DXGW con AWS Transit Gateway, varias regiones y VIF pública de AWS

Atributos del modelo de conectividad:

- La VIF pública de AWS DX se usa para acceder a recursos públicos de AWS, como S3, directamente sobre las conexiones de AWS DX.
- Proporcionar la capacidad de conectarse a VPC o conexiones de DX en otras regiones en el futuro.
- Con AWS Transit Gateway conectado a las VPC, se puede conseguir una conectividad de malla total o parcial entre ellas.
- Comunicación entre VPC y VPC entre regiones facilitada por el emparejamiento de AWS Transit Gateway.

- Ofrece opciones de diseño flexibles para integrar dispositivos virtuales de seguridad y SDWAN de terceros con AWS Transit Gateway. Consulte Seguridad de red centralizada para tráfico de VPC a VPC y en las instalaciones a VPC.

Consideraciones de escala:

- El número de rutas hacia y desde AWS Transit Gateway está limitado al número máximo admitido de rutas sobre una VIF de tránsito (los números de entrada y salida varían). Consulte las cuotas de AWS Direct Connect para obtener más información sobre los límites de escala y el número de rutas y las VIF que se admiten.
- Escale verticalmente hasta miles de VPC por AWS Transit Gateway en una única sesión BGP.
- VIF de tránsito única por AWS DX.
- Se pueden agregar conexiones de AWS DX adicionales según se desee.

Otras consideraciones:

- Genera costos adicionales de procesamiento de AWS Transit Gateway para la transferencia de datos entre AWS y el sitio en las instalaciones.
- No se puede hacer referencia a los grupos de seguridad de una VPC remota mediante AWS Transit Gateway (necesita emparejamiento de VPC).
- El emparejamiento de VPC puede utilizarse en lugar de AWS Transit Gateway para facilitar la comunicación entre las VPC. No obstante, esto agregará complejidad operativa para crear y administrar a escala un gran número de emparejamientos de VPC punto a punto.
- Si se necesitan más de tres AWS Transit Gateway, se pueden agregar DXGW adicionales; consulte el siguiente modo de conectividad.

AWS DX: DXGW con AWS Transit Gateway, varias regiones (más de tres)

Este modelo se construye a partir de:

- Varias Regiones de AWS (más de tres).
- Centros de datos dobles en las instalaciones.
- Conexiones de AWS Direct Connect dobles a ubicaciones de DX independientes por región.
- AWS DXGW con AWS Transit Gateway.
- Gran escala de VPC por región.

- Malla completa de emparejamiento entre AWS Transit Gateway.

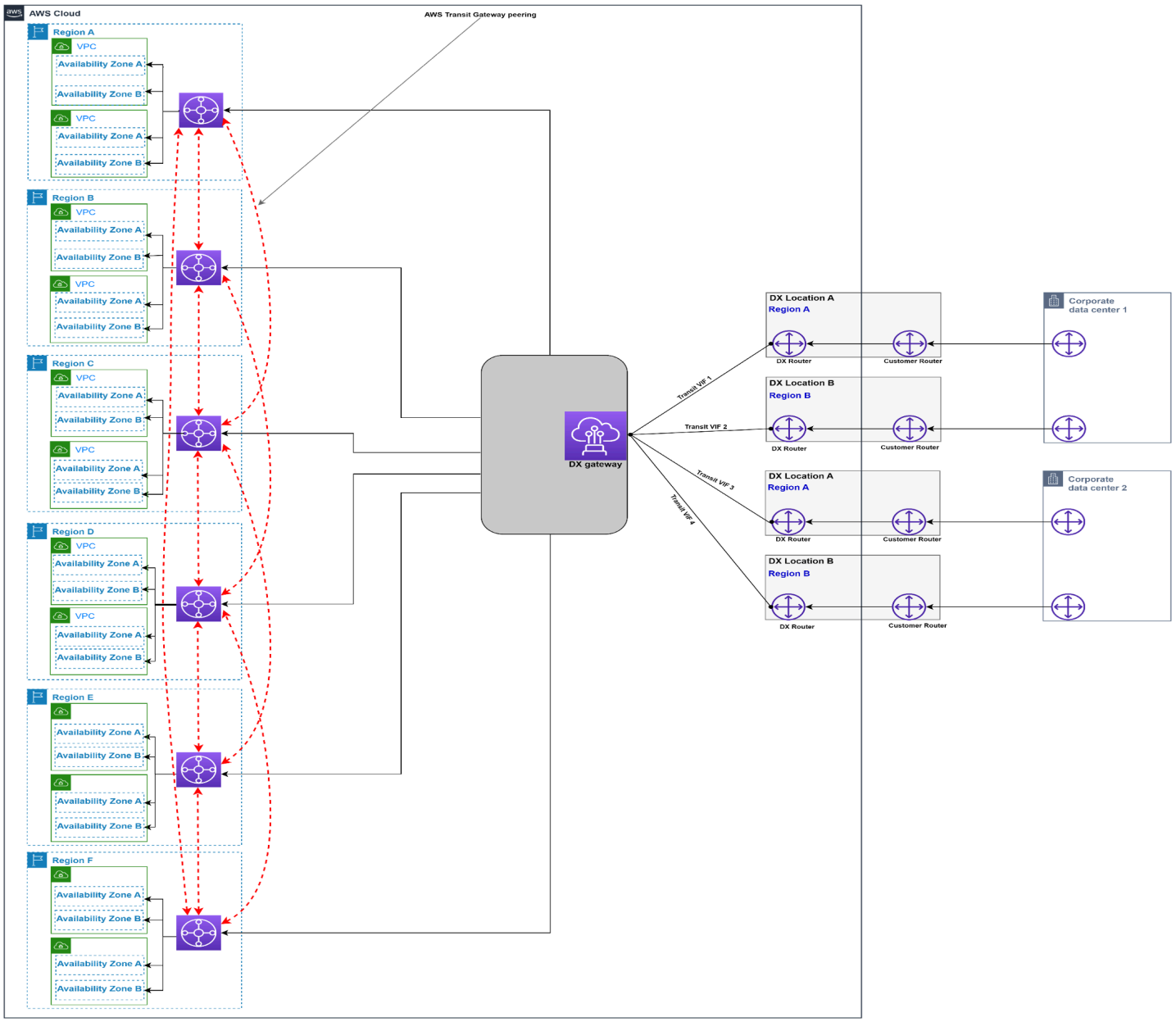

Figura 8. AWS DX: DXGW con AWS Transit Gateway, varias regiones (más de tres)

Atributos del modelo de conectividad:

- Sobrecarga operativa más baja.
- La VIF pública de AWS DX se usa para acceder a recursos públicos de AWS, como S3, directamente sobre las conexiones de AWS DX.

- Proporcionar la capacidad de conectarse a VPC y conexiones de DX en otras regiones en el futuro.
- Con AWS Transit Gateway conectado a las VPC, se puede conseguir una conectividad de malla total o parcial entre ellas.
- Comunicación de VPC entre regiones facilitada por el emparejamiento de AWS Transit Gateway.
- Ofrece opciones de diseño flexibles para integrar dispositivos virtuales de seguridad y SDWAN de terceros con AWS Transit Gateway. Consulte Seguridad de red centralizada para tráfico de VPC a VPC y en las instalaciones a VPC.

Consideraciones de escala:

- El número de rutas hacia y desde AWS Transit Gateway está limitado al número máximo admitido de rutas sobre una VIF de tránsito (los números de entrada y salida varían). Consulte las cuotas de AWS Direct Connect para obtener más información sobre los límites de escala. Considere la posibilidad de resumir las rutas si es necesario para reducir el número de rutas.
- Escalar verticalmente hasta miles de VPC por AWS Transit Gateway a través de una única sesión BGP por DXGW (suponiendo que el rendimiento proporcionado por las conexiones de AWS DX aprovisionadas sea suficiente).
- Se pueden conectar hasta seis AWS Transit Gateway por DXGW.
- Si es necesario conectar más de tres regiones mediante AWS Transit Gateway, se requieren DXGW adicionales.
- VIF de tránsito única por AWS DX.
- Se pueden agregar conexiones de AWS DX adicionales según se desee.

Otras consideraciones:

- Genera un costo adicional de procesamiento de AWS Transit Gateway para la transferencia de datos entre el sitio en las instalaciones y AWS.
- No se puede hacer referencia a los grupos de seguridad de una VPC remota mediante AWS Transit Gateway (necesita emparejamiento de VPC).
- El emparejamiento de VPC puede utilizarse en lugar de AWS Transit Gateway para facilitar la comunicación entre las VPC. No obstante, esto agregará complejidad operativa para crear y administrar a escala un gran número de emparejamientos de VPC punto a punto.

El siguiente árbol de decisiones abarca las consideraciones sobre la escalabilidad y el modelo de comunicación:

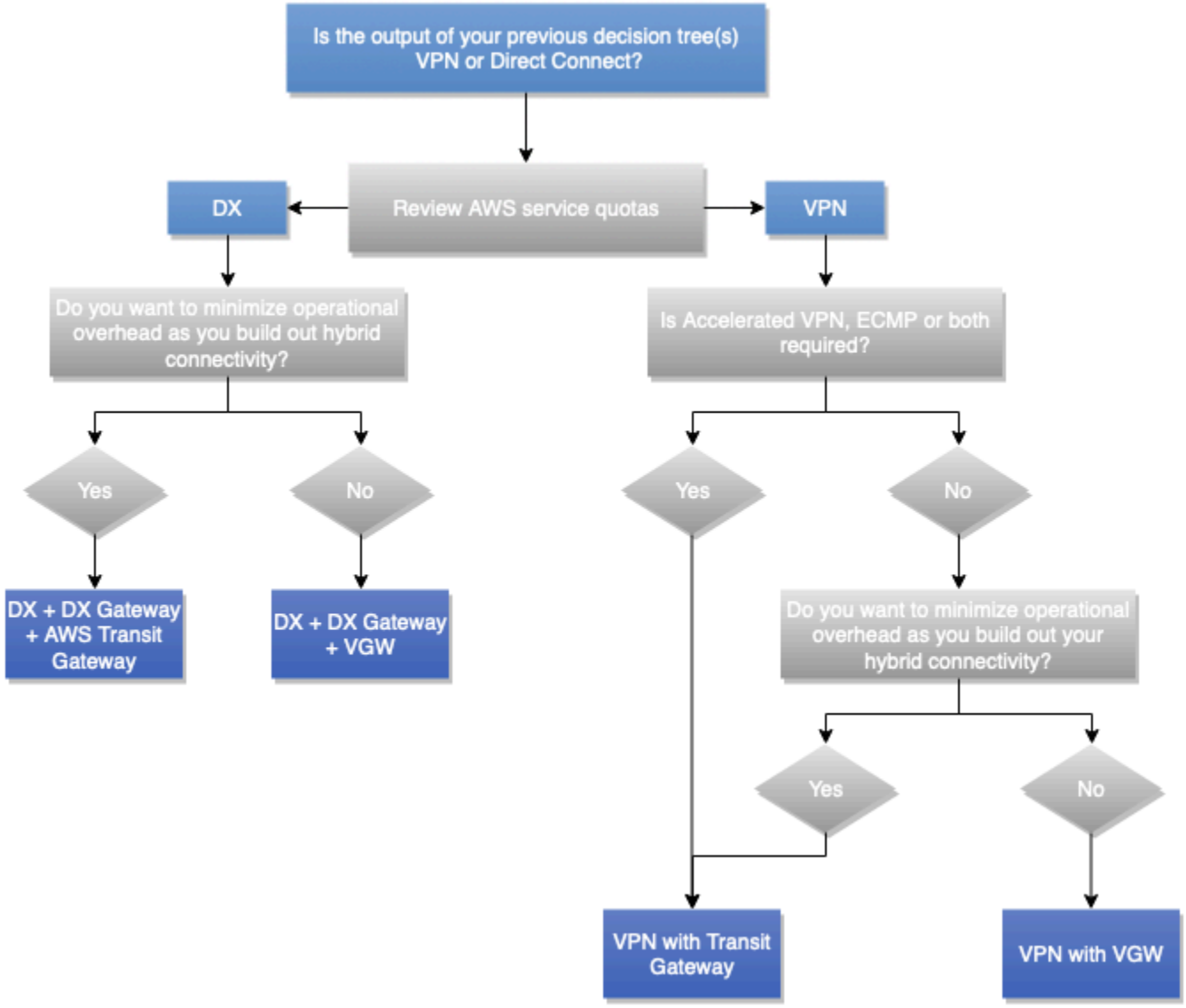

Figura 9. Árbol de decisiones del modelo de escalabilidad y comunicación

> **Note**
>
> Si el tipo de conexión seleccionado es VPN, normalmente en la consideración de rendimiento, se debe decidir si el punto de terminación de VPN es la conexión de AWS VGW o AWS Transit Gateway AWS S2S VPN. Si aún no lo ha hecho, puede tener en cuenta el modelo de comunicación necesario entre las VPC junto con el número de VPC necesarias que se conectarán a las conexiones de VPN para ayudarlo a tomar la decisión.

Fiabilidad

Definición

La fiabilidad hace referencia a la capacidad de un servicio o sistema para realizar su función prevista cuando sea necesario. La fiabilidad de un sistema puede medirse por el nivel de su calidad operativa en un plazo determinado. Se contrapone a la resiliencia, que se refiere a la capacidad de un sistema para recuperarse de las interrupciones de la infraestructura o de los servicios, de forma dinámica y fiable.

Para obtener más detalles sobre cómo se utilizan la disponibilidad y la resiliencia para medir la fiabilidad, consulte el pilar de fiabilidad del Marco de AWS Well-Architected.

Preguntas clave

Disponibilidad

La disponibilidad es el porcentaje de tiempo que una carga de trabajo está disponible para utilizarse. Entre los objetivos comunes se incluyen el 99 % (3,65 días de tiempo de inactividad permitido al año), el 99,9 % (8,77 horas) y el 99,99 % (52,6 minutos), con una abreviatura del número de nueves en el porcentaje ("dos nueves" para el 99 %, "tres nueves" para el 99,9 %, y así sucesivamente). La disponibilidad de la solución de red entre AWS y el centro de datos en las instalaciones puede ser diferente de la disponibilidad general de la solución o de la aplicación.

Entre las preguntas clave sobre la disponibilidad de una solución de red se incluyen las siguientes:

- ¿Pueden seguir funcionando mis recursos de AWS si no pueden comunicarse con mis recursos en las instalaciones? ¿Y a la inversa?
- ¿Debo considerar el tiempo de inactividad programado para el mantenimiento planificado como incluido o excluido de la métrica de disponibilidad?
- ¿Cómo mediré la disponibilidad de la capa de red, separada del estado general de la aplicación?

La sección Disponibilidad del pilar de fiabilidad del Marco de Well-Architected contiene sugerencias y fórmulas para calcular la disponibilidad.

Resistencia

La resiliencia es la capacidad de una carga de trabajo para recuperarse de interrupciones en la infraestructura o el servicio, para incorporar dinámicamente recursos computacionales que satisfagan

la demanda y para mitigar las interrupciones, como errores de configuración o problemas de red temporales. Si un componente de red redundante (enlace, dispositivos de red, etc.) no tiene la disponibilidad suficiente para proporcionar la función esperada por sí solo, significa que tiene una baja resiliencia a los errores. La consecuencia es una experiencia de usuario deficiente y degradada.

Entre las preguntas clave sobre la resiliencia de una solución de red se incluyen las siguientes:

- ¿Cuántos errores simultáneos y discretos debo permitir?
- ¿Cómo puedo reducir los puntos únicos de error tanto con las soluciones de conectividad como con mi red interna?
- ¿Cuál es mi vulnerabilidad ante los eventos de denegación de servicio distribuido (DDoS)?

Solución técnica

En primer lugar, es importante tener en cuenta que no todas las soluciones de conectividad de red híbrida requieren un elevado nivel de fiabilidad y que el aumento de los niveles de fiabilidad tiene su correspondiente incremento de costo. En algunos escenarios, un sitio principal puede requerir conexiones fiables (redundantes y resilientes), ya que el tiempo de inactividad tiene un mayor impacto en la empresa, mientras que los sitios regionales pueden no requerir el mismo nivel de fiabilidad debido al menor impacto en la empresa en caso de un evento de error. Se recomienda consultar las Recomendaciones sobre resiliencia de AWS Direct Connect, ya que explican las prácticas recomendadas de AWS para garantizar una alta resiliencia con el diseño de AWS Direct Connect.

Para lograr una solución de conectividad de red híbrida fiable en el contexto de resiliencia, el diseño debe tener en cuenta los siguientes aspectos:

- Redundancia: el objetivo es eliminar cualquier punto único de error en la ruta de conectividad de red híbrida, incluidas, entre otras, las conexiones de red, los dispositivos de red periférica, la redundancia en las zonas de disponibilidad, Regiones de AWS y las ubicaciones de DX, así como las fuentes de alimentación de los dispositivos, las rutas de fibra y los sistemas operativos. Para el propósito y el alcance de este documento técnico, la redundancia se centra en las conexiones de red, los dispositivos periféricos (por ejemplo, los dispositivos de puerta de enlace de cliente), la ubicación de AWS DX y Regiones de AWS (para arquitecturas de varias regiones).
- Componentes de conmutación por error fiables: en algunos escenarios, un sistema puede ser funcional, pero no desempeñar sus funciones al nivel requerido. Esta situación es habitual durante un evento de error único en el que se descubre que los componentes redundantes planificados

funcionaban de forma no redundante: su carga de red no tiene otro lugar al que acudir debido al uso, lo que provoca una capacidad insuficiente para toda la solución.

- Tiempo de conmutación por error: es el tiempo que tarda un componente secundario en asumir completamente el rol del componente principal. El tiempo de conmutación por error depende de varios factores: cuánto se tarda en detectar el error, cuánto se tarda en habilitar la conectividad secundaria y cuánto se tarda en notificar el cambio al resto de la red. La detección de errores puede mejorarse mediante la detección de pares muertos (DPD) en el caso de los enlaces VPN y la detección de reenvío bidireccional (BFD) en el caso de los enlaces de AWS Direct Connect. El tiempo para habilitar la conectividad secundaria puede ser muy bajo (si estas conexiones están siempre activas), puede ser un intervalo de tiempo corto (si hay que habilitar una conexión de VPN preconfigurada) o más largo (si hay que mover recursos físicos o configurar nuevos recursos). La notificación al resto de la red suele producirse a través de protocolos de enrutamiento en la red del cliente, cada uno de los cuales tiene diferentes tiempos de convergencia y opciones de configuración. En este documento técnico no se aborda la configuración de los mismos.
- Ingeniería de tráfico: en el contexto del diseño de una conectividad de red híbrida resiliente, la ingeniería de tráfico tiene como objetivo abordar cómo debe fluir el tráfico a través de las múltiples conexiones disponibles en escenarios normales y de error. Se recomienda seguir el concepto de diseño preparado para errores, en el que hay que ver cómo funcionará la solución en diferentes escenarios de error y si será aceptable para la empresa o no. En esta sección se analizan algunos de los casos de uso habituales de la ingeniería de tráfico, cuyo objetivo es mejorar el nivel general de resiliencia de la solución de conectividad de red híbrida. En la sección sobre enrutamiento y BGP de AWS Direct Connect se habla de varias opciones de ingeniería de tráfico para influir en el flujo de tráfico (comunidades, preferencia local BGP o longitud de la ruta AS). Para diseñar una solución de ingeniería de tráfico eficaz, debe conocer bien cómo cada uno de los componentes de red de AWS gestiona el enrutamiento IP en términos de evaluación y selección de rutas, así como los posibles mecanismos para influir en la selección de rutas. En este documento no se abordan los detalles al respecto. Para obtener más información, consulte la documentación Orden de evaluación de rutas de Transient Gateway, Prioridad de rutas de VPN de sitio a sitio y Enrutamiento de Direct Connect y BGP, según sea necesario.

> **Note**
>
> En la tabla de enrutamiento de VPC, puede hacer referencia a una lista de prefijos que tenga reglas adicionales de selección de rutas. Para obtener más información sobre este caso de uso, consulte la prioridad de ruta para listas de prefijos. Las tablas de enrutamiento de AWS

Transit Gateway también admiten listas de prefijos, pero, una vez aplicadas, se amplían a entradas de ruta específicas.

Ejemplo de doble conexión de VPN de sitio a sitio con rutas más específicas

Este escenario se basa en un pequeño sitio en las instalaciones que se conecta a una única región de Región de AWS mediante conexiones de VPN redundantes a través de Internet a AWS Transit Gateway. El diseño de ingeniería de tráfico representado en la figura 10 muestra que con la ingeniería de tráfico se puede influir en la selección de ruta que aumenta la fiabilidad de la solución de conectividad híbrida:

- Conectividad híbrida resiliente: las conexiones de VPN redundantes proporcionan cada una la misma capacidad de rendimiento, admiten la conmutación por error automatizada mediante el uso del protocolo de enrutamiento dinámico (BGP) y aceleran la detección de errores de conexión mediante la detección de pares muertos de VPN.
- Eficiencia de rendimiento: la configuración de ECMP en ambas conexiones de VPN a AWS Transit Gateway contribuye a maximizar el ancho de banda total de la conexión de VPN. Como alternativa, mediante el anuncio de rutas diferentes, más específicas, junto con la ruta de resumen del sitio, se puede administrar la carga de forma independiente en las dos conexiones de VPN.

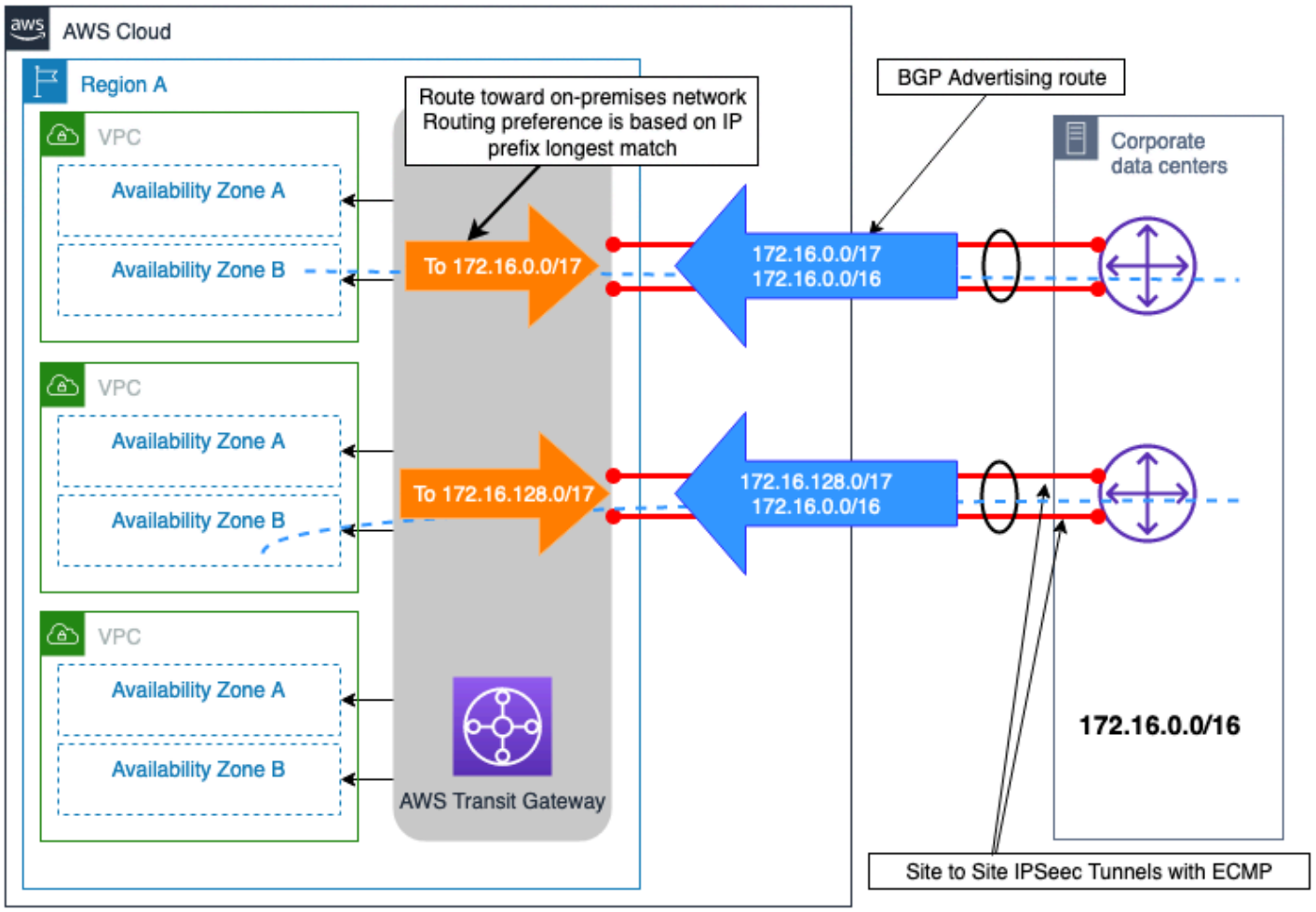

Figura 10. Ejemplo de doble conexión de VPN de sitio a sitio con rutas más específicas

Ejemplo de dos sitios en las instalaciones con varias conexiones de DX

El escenario representado en la figura 11 muestra dos sitios de centros de datos en las instalaciones situados en diferentes regiones geográficas y conectados a AWS a través del modelo de conectividad de máxima resiliencia (descrito en las Recomendaciones de resiliencia de AWS Direct Connect) mediante con AWS Direct Connect DXGW y VGW. Estos dos sitios en las instalaciones están interconectados entre sí a través de un enlace de interconexión de centros de datos (DCI). Los prefijos IP en las instalaciones (192.168.0.0/16) que pertenecen a sitios de sucursales remotas se anuncian desde ambos sitios del centro de datos en las instalaciones. La ruta principal para este prefijo debe ser el centro de datos 1. El tráfico hacia los sitios remotos de sucursales y desde ellos conmutará por error al centro de datos 2 en caso de error en el centro de datos 1 o en ambas ubicaciones de DX. Además, existe un prefijo IP específico de cada sitio para cada centro de datos.

Es necesario llegar a estos prefijos directamente y a través del otro sitio de centro de datos en caso de error de ambas ubicaciones de DX.

Al asociar atributos de comunidad BGP a las rutas anunciadas a AWS DXGW, puede influir en la selección de la ruta de salida desde AWS DXGW. Estos atributos de comunidad controlan el atributo de preferencia local BGP de AWS asignado a la ruta anunciada. Para obtener más información, consulte Políticas de enrutamiento y comunidades BGP de AWS DX.

Para maximizar la fiabilidad de la conectividad en el nivel de Región de AWS, cada par de conexiones de AWS DX configura ECMP de modo que ambas puedan utilizarse al mismo tiempo para la transferencia de datos entre cada sitio en las instalaciones y AWS.

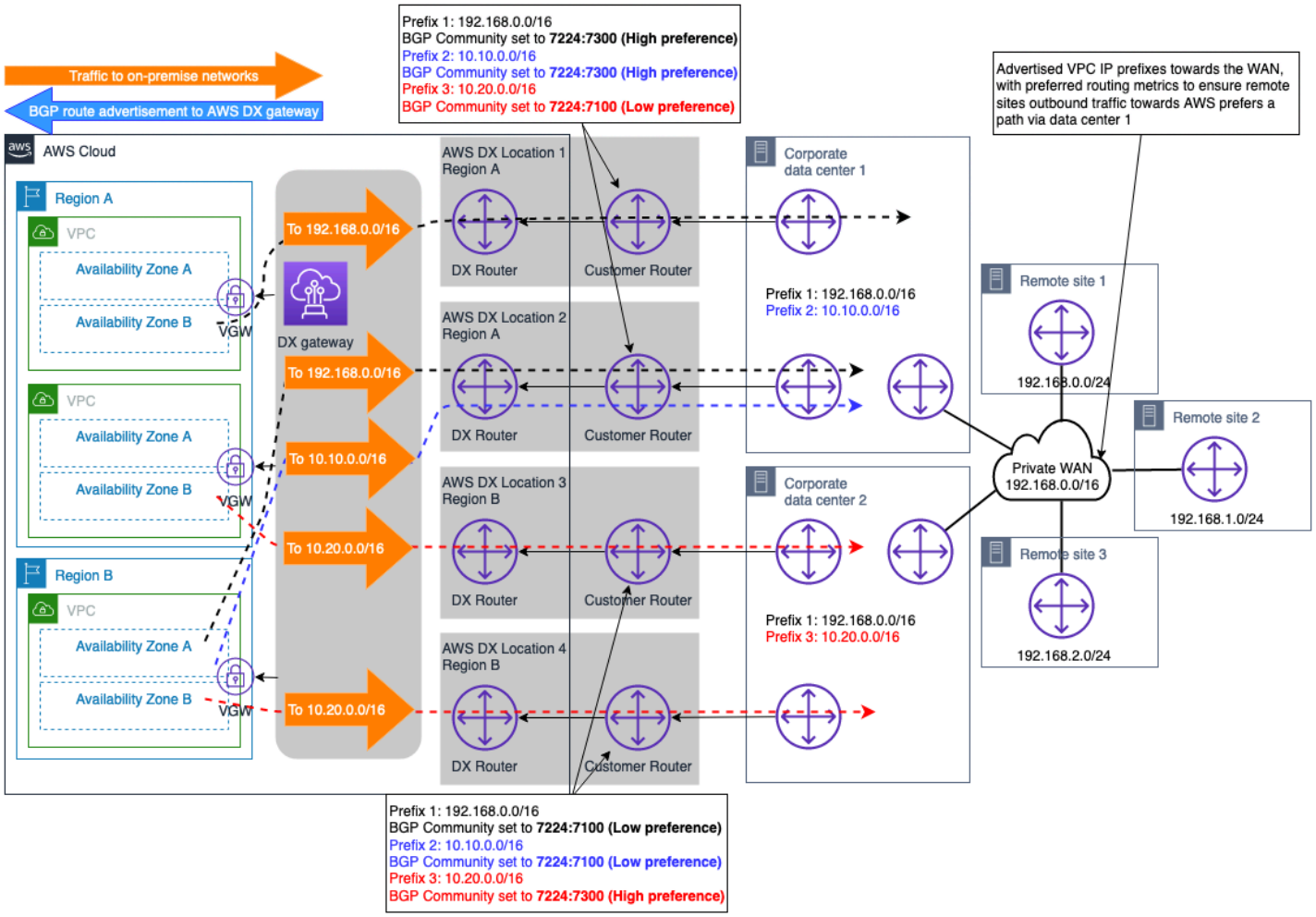

Figura 11. Ejemplo de dos sitios en las instalaciones con varias conexiones de DX

Con este diseño, los flujos de tráfico destinados a las redes en las instalaciones (con la misma longitud de prefijo anunciada y comunidad BGP) se distribuirán a través de las conexiones de DX

duales por sitio mediante ECMP. No obstante, si no se requiere ECMP a través de la conexión de DX, se puede utilizar el mismo concepto comentado anteriormente y descrito en la documentación de Routing policies and BGP communities para diseñar aún más la selección de rutas en el nivel de conexión de DX.

Nota: Si hay dispositivos de seguridad en la ruta en los centros de datos en las instalaciones, estos dispositivos deben configurarse para permitir flujos de tráfico que salgan por un enlace de DX y lleguen desde otro enlace de DX (ambos enlaces se utilizan con ECMP) en el mismo sitio de centro de datos.

Ejemplo de conexión de VPN como copia de seguridad de la conexión de AWS DX

Se puede seleccionar VPN para proporcionar una conexión de red de copia de seguridad a una conexión de AWS Direct Connect. Normalmente, este tipo de modelo de conectividad se basa en el costo, ya que proporciona un menor nivel de fiabilidad a la solución global de conectividad híbrida debido al rendimiento no determinista a través de Internet, y no existe un SLA que pueda obtenerse para una conexión a través del Internet público. Es un modelo de conectividad válido y económico, y debe utilizarse cuando el costo sea la consideración prioritaria y se disponga de un presupuesto limitado, o posiblemente como solución provisional hasta que pueda aprovisionarse un DX secundario. En la figura 12 se ilustra el diseño de este modelo de conectividad. Una consideración clave con este diseño, en el que tanto la conexión de VPN como la de DX terminan en AWS Transit Gateway, es que la conexión de VPN puede anunciar un mayor número de rutas en comparación con las que se pueden anunciar a través de una conexión de DX realizada a AWS Transit Gateway. Esto puede provocar una situación de enrutamiento poco óptima. Una opción para resolver este problema es configurar el filtrado de rutas en el dispositivo de puerta de enlace de cliente (CGW) para las rutas recibidas de la conexión de VPN, de forma que solo se permitan las rutas de resumen.

Nota: Para crear la ruta resumen en AWS Transit Gateway, es necesario especificar una ruta estática a un adjunto arbitrario en la tabla de enrutamiento de AWS Transit Gateway para que el resumen se envíe a lo largo de la ruta más específica.

Desde el punto de vista de la tabla de enrutamiento de AWS Transit Gateway, las rutas para el prefijo en las instalaciones se reciben tanto de la conexión de AWS DX (a través de DXGW) como de VPN, con la misma longitud de prefijo. Siguiendo la lógica de prioridad de rutas de AWS Transit Gateway, las rutas recibidas a través de Direct Connect tienen una mayor preferencia que las recibidas a través de VPN de sitio a sitio y, por lo tanto, la ruta a través de AWS Direct Connect será la preferida para llegar a las redes en las instalaciones.

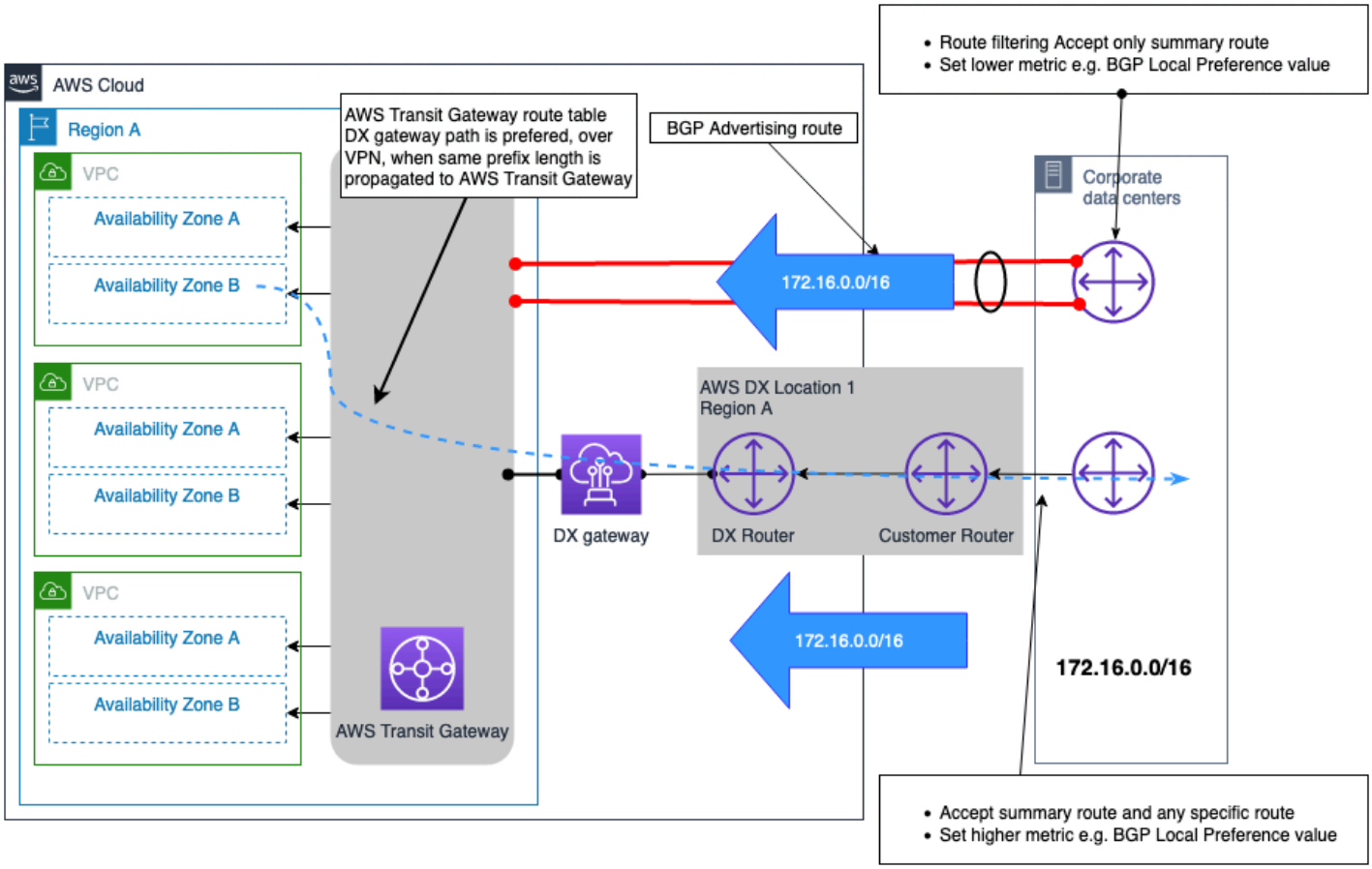

Figura 12. Ejemplo de conexión de VPN como copia de seguridad de la conexión de AWS DX

El siguiente árbol de decisiones le servirá de guía a la hora de tomar la decisión deseada para conseguir una conectividad de red híbrida resiliente (lo que se traducirá en una red fiable). Para obtener más información, consulte Kit de herramientas de resiliencia de AWS Direct Connect.

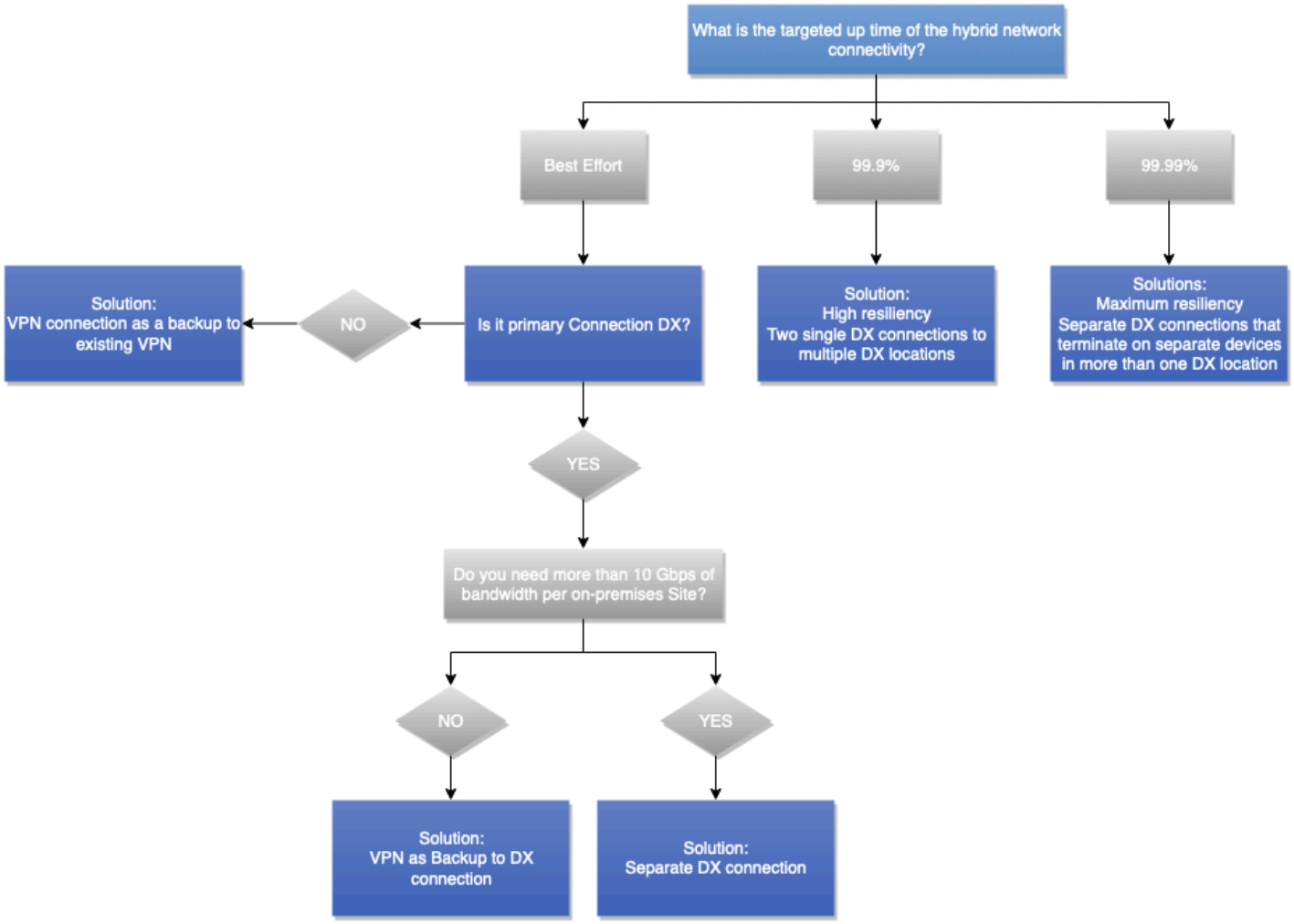

Figura 13. Árbol de decisiones de fiabilidad

VPN y SD-WAN administradas por el cliente

Definición

La conectividad a Internet es un producto básico y el ancho de banda disponible sigue aumentando cada año. Algunos clientes eligen crear una WAN virtual sobre Internet en lugar de crear y administrar una WAN privada. Una red de área extensa definida por software (SD-WAN) permite a las empresas aprovisionar y administrar de forma centralizada y rápida esta WAN virtual mediante un uso inteligente del software. Otros clientes deciden adoptar VPN tradicionales de sitio a sitio autoadministradas.

Impacto en las decisiones de diseño

La SD-WAN y las VPN administradas por el cliente pueden funcionar a través de Internet o AWS Direct Connect. SD-WAN (o cualquier superposición de VPN por software) es tan fiable como el transporte de red subyacente. Por lo tanto, las consideraciones sobre fiabilidad y SLA comentadas anteriormente en este documento técnico son aplicables en este caso. Por ejemplo, crear una superposición de SD-WAN por Internet no ofrecerá la misma fiabilidad que si se crea sobre AWS Direct Connect.

Definición de requisitos

- ¿Utiliza SD-WAN en su red en las instalaciones?
- ¿Hay características específicas que necesite y que solo estén disponibles en determinados dispositivos virtuales utilizados para la terminación de la VPN?

Soluciones técnicas

AWS recomienda integrar SD-WAN con AWS Transit Gateway y publica una lista de los proveedores que admiten la integración con AWS Transit Gateway. AWS puede actuar como concentrador de sitios SD-WAN o como sitio radial. La red troncal de AWS puede utilizarse para conectar diferentes concentradores SD-WAN implementados en AWS con una red de gran fiabilidad y rendimiento. Las soluciones SD-WAN admiten la conmutación por error automatizada a través de cualquier ruta disponible, monitorización adicional y capacidades de observabilidad en un único panel de administración. El amplio uso de la autoconfiguración y la automatización permite un aprovisionamiento y una visibilidad rápidos en comparación con las WAN tradicionales. No obstante, el uso de túneles y la sobrecarga de cifrado no son comparables a los enlaces de fibra dedicados y de alta velocidad utilizados en la conectividad privada.

En algunos casos, puede elegir utilizar un dispositivo virtual con capacidad de VPN. Entre las razones para elegir un dispositivo virtual autoadministrado figuran las características técnicas y la compatibilidad con el resto de la red. Cuando seleccione una VPN autoadministrada o una solución SD-WAN que utilice un dispositivo virtual implementado en una instancia de EC2, será responsable de la administración de dicho dispositivo. También es responsable de la alta disponibilidad y la conmutación por error entre dispositivos virtuales. Este diseño aumenta su responsabilidad operativa; no obstante, podría proporcionarle más flexibilidad. Las características y capacidades de la solución dependen del dispositivo virtual que seleccione.

AWS Marketplace contiene muchos dispositivos virtuales de VPN que los clientes pueden implementar en Amazon EC2. AWS recomienda comenzar con S2S VPN administrado AWS y buscar otras opciones si no satisface sus necesidades. La sobrecarga de administración de los dispositivos virtuales es responsabilidad del cliente.

Caso de uso de Example Corp. Automotive

En esta sección del documento técnico se muestra cómo se utilizan las consideraciones, las preguntas de definición de requisitos y los árboles de decisiones para ayudarlo a decidir el diseño de red híbrida óptimo. Es importante identificar y capturar los requisitos, ya que se utilizan como información para los árboles de decisiones. Al capturar los requisitos por adelantado, se evitan iteraciones de diseño posteriores. Detener un proyecto por completo si hay que revisar el diseño y tener recursos valiosos en espera puede minimizarse e, idealmente, evitarse cuando los requisitos se conocen por adelantado.

En esta sección se utilizará Example Corp. Automotive como cliente ilustrativo. Desea implementar inicialmente su primer proyecto de análisis en AWS. El proyecto de análisis se centra en analizar los datos de los automóviles fabricados por la empresa y otros conjuntos de datos que ya existen en los centros de datos de la empresa. Inicialmente, el grupo de arquitectura de la empresa cree que necesitará una Cuenta de AWS, una Amazon VPC y unas cuantas subredes para alojar los entornos de producción y desarrollo. El equipo de proyecto está impaciente por empezar y ha solicitado acceso al entorno de desarrollo lo antes posible. Su objetivo es pasar a producción dentro de tres meses.

Example Corp. Automotive también tiene previsto utilizar AWS para varios proyectos adicionales, como la migración de sus sistemas ERP, la infraestructura de escritorios virtuales (VDI) y otras 20 aplicaciones desde las instalaciones a AWS en los próximos 6 meses. Todavía se están definiendo algunos requisitos para proyectos adicionales, pero está claro que su uso de Nube de AWS va a aumentar.

El equipo de arquitectura decidió aprovechar el enfoque descrito en este documento técnico. Utilizó las preguntas de definición de requisitos esbozadas en cada consideración para captar las aportaciones que lo permitieran tomar sus decisiones de diseño.

Comienza con los requisitos relacionados con el tipo de conectividad que se resumen en la siguiente tabla.

Table 4. Entradas de fiabilidad de Example Automotive Corp

Consideraciones sobre la selección del tipo de conectividad	Preguntas sobre la definición de requisitos	Respuestas
Momento de implementación	¿Cuál es el plazo requerido para la implementación? ¿Horas, días, semanas o meses?	• Desarrollo/pruebas: 1 mes • Producción: 3 meses
Seguridad	¿Permiten sus requisitos y políticas de seguridad utilizar conexiones cifradas a través de Internet para conectarse a AWS u obligan a utilizar conexiones de red privada?	• Desarrollo/pruebas: se acepta VPN de sitio a sitio • Producción: se requiere una red privada
	Cuando se aprovechan las conexiones de red privada, ¿la capa de red tiene que proporcionar cifrado en tránsito?	No, se utilizará el cifrado de la capa de aplicación.
SLA	¿Se requiere un SLA de conectividad híbrida con créditos de servicio?	• Desarrollo/pruebas: no • Producción: sí
	¿Cuál es el objetivo de tiempo de actividad?	• Desarrollo/pruebas: N/A • Producción: 99,99 %
	¿Cumple toda la red híbrida el objetivo de tiempo de actividad?	• Desarrollo/pruebas: N/A • Producción: sí
Rendimiento	¿Cuál es el rendimiento requerido?	• Desarrollo/pruebas: 100 Mbps • Producción: 500 Mbps hasta 2 Gbps

Consideraciones sobre la selección del tipo de conectividad	Preguntas sobre la definición de requisitos	Respuestas
	¿Cuál es la latencia máxima aceptable entre AWS y la red en las instalaciones?	• Desarrollo/pruebas: sin requisitos estrictos • Producción:menos de 30 ms
	¿Cuál es la fluctuación de red máxima aceptable?	• Desarrollo/pruebas: sin requisitos estrictos • Producción: se requiere una fluctuación mínima
Costo	¿Cuántos datos enviaría a AWS al mes?	• Desarrollo/pruebas: 2 TB • Producción: 20 TB hasta 50 TB
	¿Cuántos datos enviaría desde AWS al mes?	• Desarrollo/pruebas: 1 TB • Producción: 10 TB hasta 25 TB
	¿Esta conectividad es permanente?	Sí

En función de los requisitos recibidos, el equipo de arquitectura siguió el árbol de decisiones del tipo de conectividad de la figura 9. Permitió al equipo de arquitectura decidir el tipo de conectividad para el desarrollo y los entornos de prueba y producción. Para el entorno de producción, tuvo en cuenta tanto los requisitos inmediatos como los futuros. Para el desarrollo y las pruebas Example Corp. Automotive establecerá una VPN de sitio a sitio a través de Internet. Para la producción, va a trabajar con un proveedor de servicios para conectar su red corporativa con AWS Direct Connect. Example Corp. Automotive consideró inicialmente utilizar una conexión alojada de Direct Connect. No obstante, debido a los requisitos de un SLA proporcionado por AWS, seleccionaron las conexiones dedicadas de Direct Connect.

Tras decidir el tipo de conectividad, el siguiente paso consiste en captar los requisitos que influyen en la selección del diseño de conectividad. Esto está relacionado con el diseño lógico, por ejemplo,

cómo se configuran las conexiones y qué servicios de AWS utilizar para respaldar los requisitos empresariales y técnicos.

Para captar los requisitos de escalabilidad y del modelo de comunicación, el equipo de arquitectura utilizó las preguntas de definición de requisitos de las secciones asociadas de este documento técnico. Los requisitos relacionados con esas dos consideraciones se resumen en la siguiente tabla.

Tabla 5. Preguntas sobre la definición de requisitos

Consideraciones sobre la selección del diseño de conectividad	Preguntas sobre la definición de requisitos	Respuestas
Escalabilidad	¿Cuál es el número actual o previsto de VPC que requieren conectividad con los sitios en las instalaciones?	2 inicialmente, hasta 30 en 6 meses
	¿Se implementan estas VPC en una única Región de AWS o en varias?	Región única
	¿Cuántos sitios en las instalaciones deben conectarse a AWS?	2 centros de datos
	¿Cuántos dispositivos de puerta de enlace de cliente tiene por sitio que necesita conectar a AWS?	2 enrutadores por centro de datos
	¿Cuántas rutas se espera que se anuncien a AWS VPC, así como el número de rutas que se espera recibir desde AWS?	• Rutas que se anunciarán a AWS: 20 rutas • Rutas que se recibirán de AWS: 1 ruta /16
	¿Hay algún plan para tener en cuenta el aumento del ancho	• Desarrollo/pruebas: 100 Mbps

Consideraciones sobre la selección del diseño de conectividad	Preguntas sobre la definición de requisitos	Respuestas
	de banda de la conexión a AWS en un futuro próximo?	• Producción: 500 Mbps hasta 2 Gbps.
Modelos de diseño de conectividad	¿Es necesario habilitar la comunicación entre VPC (en una región o entre regiones)?	Sí, en una Región de AWS
	¿Existe un requisito para acceder a los servicios de puntos de conexión públicos de AWS directamente desde las instalaciones?	Sí
	¿Existe algún requisito para acceder a los servicios de AWS mediante puntos de conexión de VPC desde las instalaciones?	No

Basándose en las aportaciones, el equipo de arquitectura siguió el árbol de decisiones de la sección Diseño de conectividad. Tras prever que el número de VPC va a crecer de 2 a 30 en los próximos 6 meses, el equipo de arquitectura decidió utilizar AWS Transit Gateway como puerta de enlace de terminación de la conexión y el enrutamiento entre VPC. AWS Transit Gateways independientes terminarán la conexión de VPN utilizada para el desarrollo y las pruebas, y para la conectividad de producción con AWS Direct Connect. El uso de AWS Transit Gateway independientes simplifica la administración de los cambios y proporciona una demarcación clara entre los entornos de desarrollo/pruebas y producción. Para la producción, se requiere puerta de enlace de AWS Direct Connect debido a AWS Transit Gateway. Se utilizará una VIF pública para acceder a los servicios de puntos de conexión públicos de AWS. En la figura 14 se ilustra el recorrido realizado en el árbol de decisiones a partir de los requisitos recopilados.

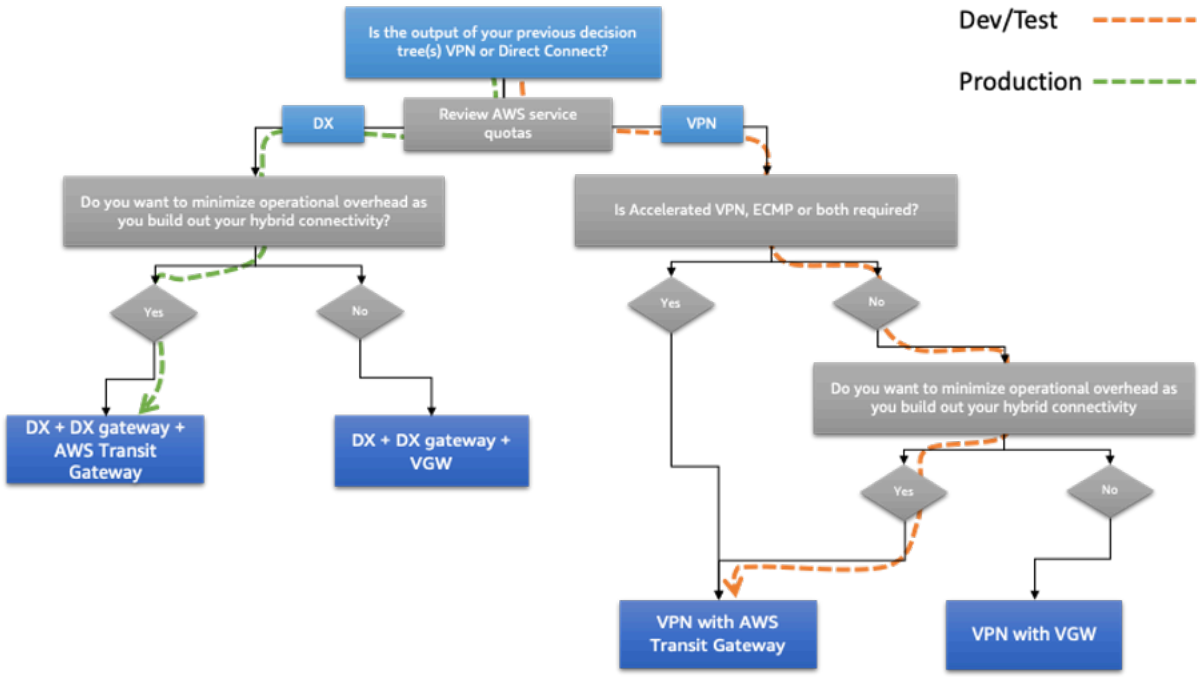

Figura 14. Árbol de decisiones de diseño de conexiones de Example Corp. Automotive

Una vez decidida la solución para cumplir los requisitos de escalabilidad y modelo de comunicación, el siguiente paso es captar los requisitos asociados a la fiabilidad. Esto está relacionado con el nivel requerido de disponibilidad y resiliencia.

Para captar los requisitos de fiabilidad, el equipo de arquitectura utilizó las preguntas de definición de requisitos de la sección asociada de este documento técnico. Los requisitos se resumen en la siguiente tabla.

Tabla 6. Preguntas sobre los requisitos de fiabilidad

Consideraciones sobre la selección del diseño de conectividad	Preguntas sobre la definición de requisitos	Respuestas
Fiabilidad	¿Cuál es la magnitud del impacto en la empresa en caso de un error de conectividad a AWS?	• Desarrollo/pruebas: bajo • Producción: alto
	Desde un punto de vista empresarial, ¿el costo de	• Desarrollo/pruebas: no

Consideraciones sobre la selección del diseño de conectividad	Preguntas sobre la definición de requisitos	Respuestas
	seguir un error de conectividad a AWS compensa el costo de implementar un modelo de conectividad de alta fiabilidad a AWS?	• Producción: sí

Basándose en las aportaciones recibidas, el equipo de arquitectura siguió el árbol de decisiones de las secciones sobre consideraciones de fiabilidad tratadas anteriormente en este documento técnico. Tras considerar el objetivo de tiempo de actividad del 99,99 % para la conectividad de producción y el elevado impacto empresarial si se producía una interrupción del servicio, el equipo de arquitectura decidió utilizar dos ubicaciones de Direct Connect y disponer de dos enlaces desde cada centro de datos en las instalaciones a cada ubicación de Direct Connect (cuatro enlaces en total). La conexión de VPN utilizada para el desarrollo y las pruebas también utilizará dos conexiones de VPN para una redundancia adicional. Mediante las técnicas de ingeniería de rutas comentadas en la sección de fiabilidad, la conectividad se configurará de la siguiente manera:

- Para el desarrollo y las pruebas, el tráfico va a tener una carga equilibrada mediante ECMP a través de los dos túneles que van al centro de datos principal. Esto permite un mayor rendimiento. Los túneles que van al centro de datos secundario se van a utilizar en caso de error de los túneles principales.
- Para la producción, la latencia entre las instalaciones y AWS a través de cualquiera de las ubicaciones de Direct Connect es muy similar. En este caso, se ha decidido equilibrar la carga del tráfico entre AWS y en las instalaciones a través de las dos conexiones que van al centro de datos principal para los sistemas en las instalaciones implementados en el centro de datos principal. Del mismo modo, para los sistemas en las instalaciones que se ejecutan en el centro de datos secundario, el tráfico va a tener una carga equilibrada entre las dos conexiones al centro de datos secundario. En caso de error en las conexiones, BGP facilitará una conmutación por error automatizada.

En la figura 15 se ilustra el recorrido realizado en el árbol de decisiones a partir de los requisitos recopilados.

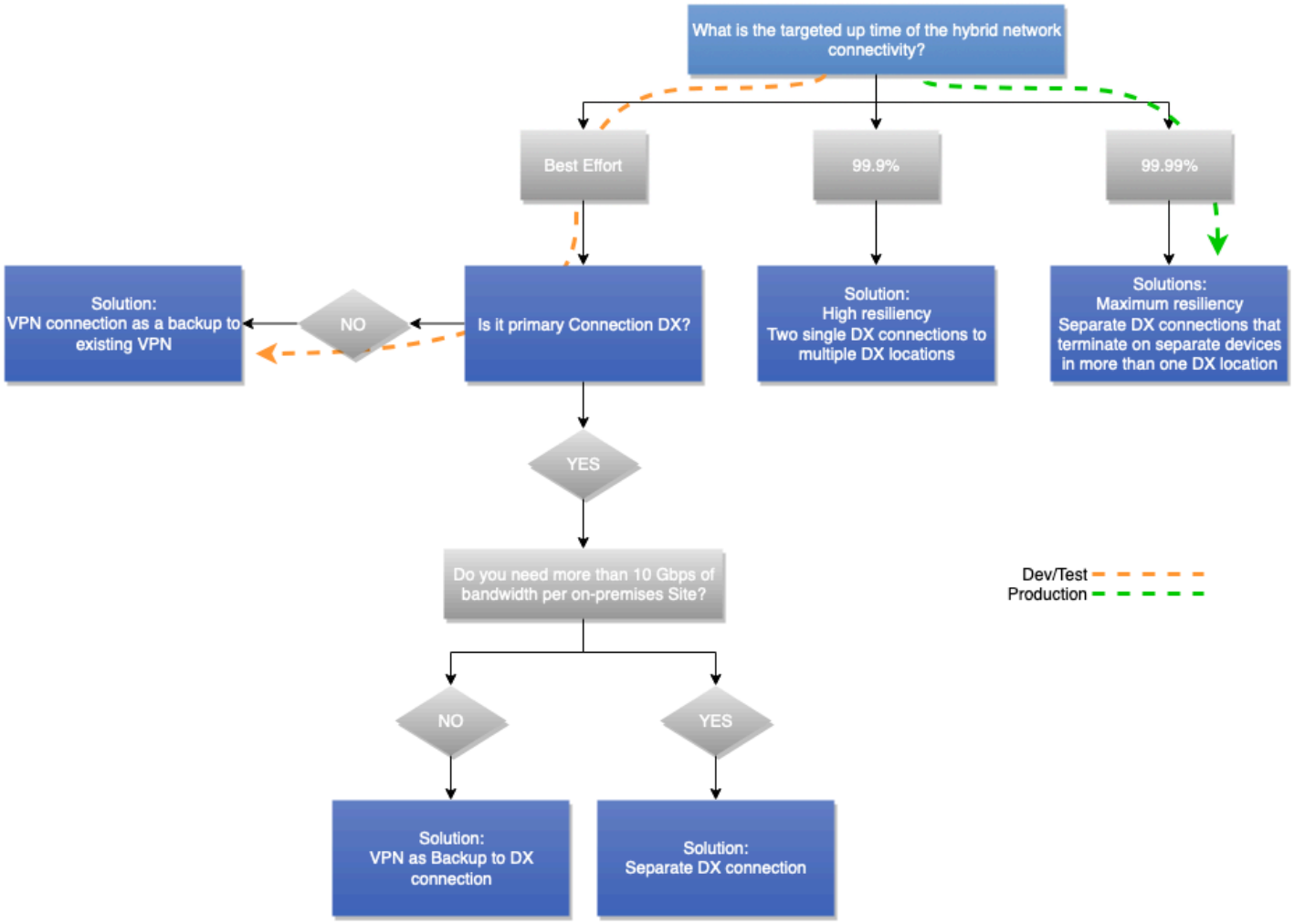

Figura 15. Árbol de decisiones de fiabilidad de Example Corp. Automotive

Arquitectura seleccionada por Example Corp. Automotive

En el siguiente diagrama se ilustra la arquitectura seleccionada por Example Corp. Automotive tras recopilar los requisitos y navegar por los árboles de decisiones tratados en las secciones anteriores de este documento técnico.

Utiliza AWS S2S VPN a través de Internet que termina en AWS Transit Gateway para el desarrollo y las pruebas. A continuación, utiliza AWS Direct Connect con puerta de enlace de Direct Connect y un segundo AWS Transit Gateway para el tráfico de producción. AWS Transit Gateway se utiliza para el enrutamiento entre VPC. Desde la perspectiva de las rutas de datos, los túneles VPN para el centro de datos principal se utilizan como rutas principales para el desarrollo y las pruebas, y los túneles hacia el centro de datos secundario se utilizan como rutas de conmutación por error. Para el

tráfico de producción, todas las conexiones se utilizan simultáneamente. El tráfico de AWS prefiere la conexión de red más opcional en función del centro de datos en el que se encuentre el sistema en las instalaciones. Example Corp. Automotive utiliza técnicas similares de ingeniería de rutas para preferir la ruta adecuada cuando el tráfico se envía a AWS, lo que garantiza que se utilicen rutas de tráfico simétricas para minimizar el uso de la red corporativa entre los centros de datos principal y secundario en las instalaciones.

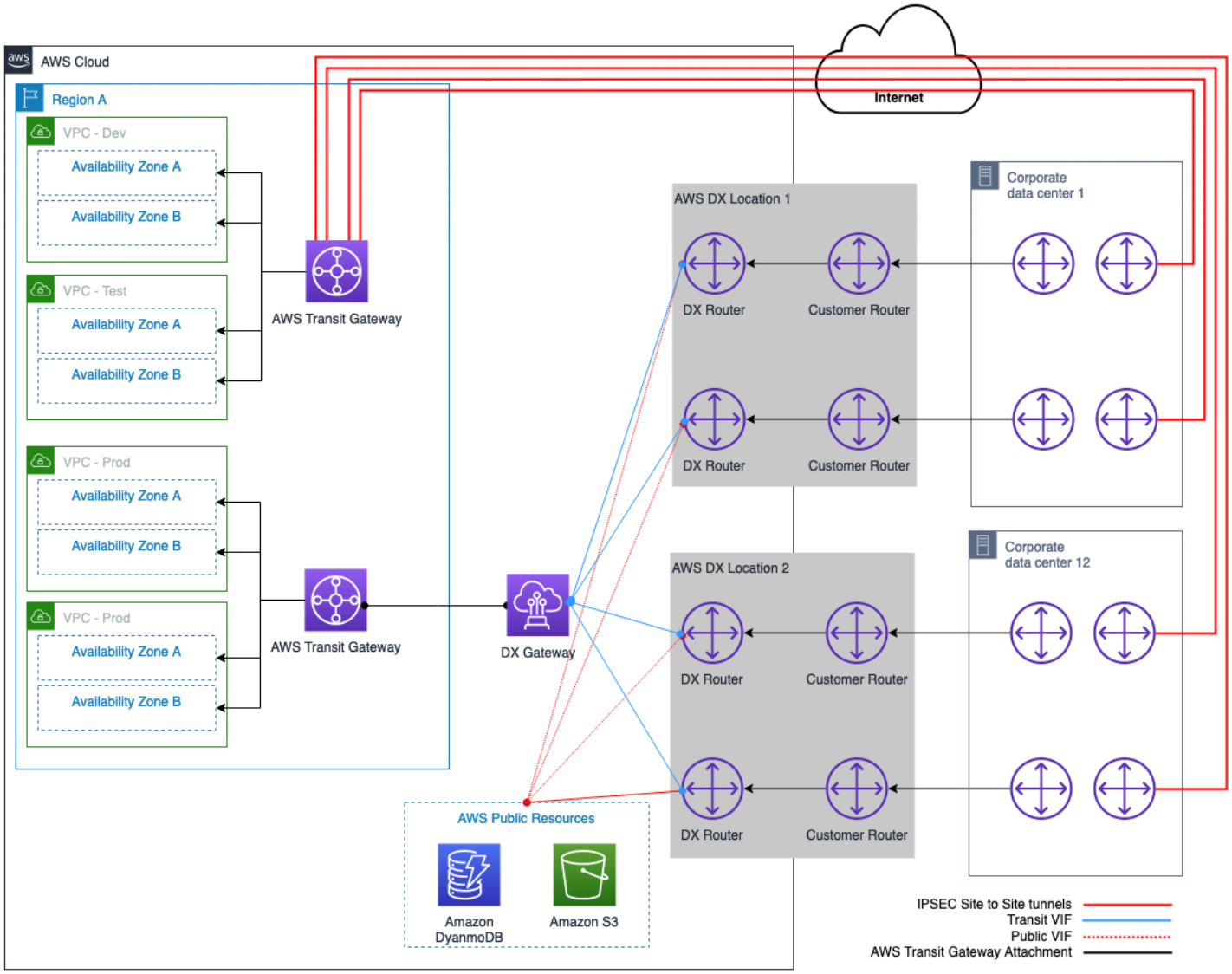

Figura 16: Example Corp. Automotive seleccionó un modelo de conectividad híbrida

Conclusión

Un modelo de conectividad híbrida es uno de los puntos de partida fundamentales para la adopción de la computación en la nube. Se puede crear una red híbrida con una arquitectura óptima según el proceso de selección del modelo de conectividad descrito en este documento técnico.

El proceso consta de consideraciones dispuestas en un orden lógico. El orden se asemeja mucho a un modelo mental que siguen los arquitectos experimentados en redes y nubes. En cada grupo de consideraciones, los árboles de decisiones permiten una selección rápida del modelo de conectividad, incluso con requisitos de entrada limitados. Puede que algunas consideraciones y sus correspondientes impactos apunten a soluciones distintas. En esos casos, como responsable de la toma de decisiones, es posible que deba ceder en algunos requisitos y seleccionar la solución más óptima que satisfaga sus necesidades empresariales y técnicas.

Colaboradores

Los colaboradores de este documento son:

- James Devine, arquitecto principal de soluciones, Amazon Web Services
- Andrew Gray, arquitecto principal de soluciones - redes, Amazon Web Services
- Maks Khomutskyi, arquitecto sénior de soluciones, Amazon Web Services
- Marwan Al Shawi, arquitecto de soluciones, Amazon Web Services
- Santiago Freitas, director de tecnología, Amazon Web Services
- Evgeny Vaganov, arquitecto especialista en soluciones - redes, Amazon Web Services
- Tom Adamski, arquitecto especialista en soluciones - redes, Amazon Web Services

Documentación adicional

- Creación de una infraestructura de red de AWS multiVPC escalable y segura
- Opciones de DNS en la nube híbrida para Amazon VPC
- Opciones de conectividad de Amazon Virtual Private Cloud
- Documentación sobre Amazon Virtual Private Cloud
- Documentación de AWS Direct Connect
- ¿Cuál es la diferencia entre una interfaz virtual alojada (VIF) y una conexión alojada?

Revisiones del documento

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

Cambio	Descripción	Fecha
Actualización menor	Se actualizó para reflejar el aumento del límite de cuota de DX.	10 de julio de 2023
Actualización importante	Se actualizó para incorporar las últimas prácticas recomendadas, servicios y capacidades.	6 de julio de 2023
Actualización menor	Se actualizaron los diagramas de arquitectura de referencia para reflejar los cambios en la cuota de DX.	27 de junio de 2023
Actualización menor	Se corrigieron los enlaces que no funcionaban.	22 de marzo de 2022
Publicación inicial	Documento técnico publicado por primera vez	22 de septiembre de 2020

Avisos

Es responsabilidad de los clientes realizar su propia evaluación independiente de la información que contiene este documento. El presente documento: (a) tiene solo fines informativos, (b) representa las ofertas y prácticas actuales de los productos de AWS, que están sujetas a cambios sin previo aviso, y (c) no supone ningún compromiso ni garantía por parte de AWS ni sus filiales, proveedores o licenciantes. Los productos o servicios de AWS se proporcionan “tal cual”, sin garantías, declaraciones ni condiciones de ningún tipo, ya sean expresas o implícitas. Las responsabilidades y obligaciones de AWS con respecto a sus clientes se controlan mediante los acuerdos de AWS y este documento no forma parte ni modifica ningún acuerdo entre AWS y sus clientes.

© 2023 Amazon Web Services, Inc. o sus filiales. Todos los derechos reservados.

Glosario de AWS

Para ver la terminología más reciente de AWS, consulte el Glosario de AWS en la Referencia de Glosario de AWS.