AWS Livro branco

Conectividade híbrida

Conectividade híbrida: AWS Livro branco

Table of Contents

Conectividade híbrida

Data de publicação: 16 de julho de 2023 (Revisões do documento)

Muitas organizações precisam conectar seus datacenters on-premises, locais remotos e a nuvem. Uma rede híbrida conecta esses diferentes ambientes. Este whitepaper descreve os componentes básicos da AWS e os principais requisitos a serem considerados ao decidir qual modelo de conectividade híbrida é ideal para você. Para ajudá-lo a determinar a melhor solução para seus requisitos comerciais e técnicos, fornecemos árvores de decisão para guiá-lo pelo processo lógico de seleção.

Introdução

Uma organização moderna usa uma ampla variedade de recursos de TI. No passado, era comum hospedar esses recursos em um datacenter on-premises ou em uma instalação de colocalização. Com o aumento da adoção da computação em nuvem, as organizações fornecem e consomem recursos de TI de provedores de serviços em nuvem por meio de uma conexão de rede. As organizações podem optar por migrar alguns ou todos os seus recursos de TI existentes para a nuvem. Em ambos os casos, é necessária uma rede comum para conectar recursos on-premises e na nuvem. A coexistência de recursos locais e na nuvem é chamada de nuvem híbrida, e a rede comum que os conecta é chamada de rede híbrida. Mesmo que sua organização mantenha todos os seus recursos de TI na nuvem, ela ainda pode precisar de conectividade híbrida com locais remotos.

Existem vários modelos de conectividade para você escolher. Embora ter opções aumente a flexibilidade, selecionar a opção ideal exige a análise dos requisitos comerciais e técnicos e a eliminação das opções que não são adequadas. Você pode agrupar os requisitos em considerações como segurança, tempo para implantação, desempenho, confiabilidade, modelo de comunicação, escalabilidade e muito mais. Depois de coletar, analisar e considerar cuidadosamente os requisitos, os arquitetos de rede e nuvem podem identificar os componentes e as soluções de rede híbrida da AWS aplicáveis. Para identificar e selecionar o modelo ou modelos ideais, os arquitetos devem compreender as vantagens e desvantagens de cada modelo. Também existem limitações técnicas que podem fazer com que um modelo adequado seja excluído.

Para simplificar o processo de seleção, este whitepaper orienta você em cada consideração importante em uma ordem lógica. Em cada consideração, há perguntas usadas para coletar requisitos. O impacto de cada decisão de projeto é identificado, junto com as possíveis soluções. O whitepaper apresenta árvores de decisão para algumas das considerações como um método para

auxiliar o processo de tomada de decisão, eliminar opções e compreender as consequências de cada decisão. Ele conclui com um cenário que abrange um caso de uso híbrido, aplicando a seleção e o design do modelo de conectividade de ponta a ponta. Você pode usar esse exemplo para ver como executar os processos apresentados neste whitepaper em um exemplo prático.

O objetivo deste whitepaper é ajudar você a selecionar e projetar um modelo de conectividade híbrida ideal. Este whitepaper é estruturado da maneira a seguir:

- Blocos de construção da conectividade híbrida – Uma visão geral dos serviços da AWS usados para conectividade híbrida.
- Seleção de conectividade e considerações de design – Uma definição de cada modelo de conectividade, como cada um afeta a decisão de design, questões de identificação de requisitos, soluções e árvores de decisão.
- Um caso de uso do cliente - Um exemplo de como aplicar as considerações e as árvores de decisão na prática.

Você é Well-Architected?

O AWSWell-Architected Framework ajuda você a entender os prós e os contras das decisões que você toma ao criar sistemas na nuvem. Os seis pilares do framework permitem a você conhecer as melhores práticas de arquitetura para criar e operar sistemas confiáveis, seguros, econômicos e sustentáveis na nuvem. Usando o AWS Well-Architected Tool, disponível gratuitamente no AWS Management Console, você pode analisar suas cargas de trabalho em relação a essas melhores práticas respondendo a um conjunto de perguntas para cada pilar.

Para obter orientações especializadas e melhores práticas adicionais para a arquitetura de sua nuvem (implantações de arquitetura de referência, diagramas e whitepapers), consulte o AWS Architecture Center.

Blocos de construção de conectividade híbrida da AWS

Há três blocos básicos de construção uma arquitetura de conectividade de rede híbrida:

- Conexões de rede híbrida: os tipos de conexão entre serviços de conectividade da AWS e dispositivos de gateway do cliente on-premises.
- Serviços de conectividade híbrida da AWS: os serviços da AWS que fornecem conectividade e roteamento entre a infraestrutura do cliente e a AWS.
- Dispositivo de gateway do cliente on-premises: o dispositivo dentro da rede existente do cliente que é o endpoint on-premises para conexão de rede híbrida. Diferentes tipos de conexão têm requisitos técnicos diferentes para esses dispositivos, que são discutidos nas seções a seguir.

Conexões de rede híbridas

Existem várias maneiras de fazer a conexão entre seus equipamentos on-premises e a AWS. Este whitepaper se concentra em como essas diferentes maneiras podem ser combinadas em arquiteturas gerais. No entanto, é fornecida uma breve visão geral das diferentes opções (AWS Direct Connect, Site-to-Site Virtual Private Network e Transit Gateway Connect).

AWS Direct Connect

O AWS Direct Connect é um serviço que estabelece uma conexão de rede dedicada de suas instalações para a AWS. Para mais detalhes, consulte AWS Direct Connect.

Há dois tipos de conexões do AWS Direct Connect: dedicadas e hospedadas. Uma conexão dedicada é um link direto entre um dispositivo AWS e seu dispositivo on-premises, enquanto uma conexão hospedada é fornecida por um parceiro da AWS que pode lidar com os detalhes da conexão para você. Consulte Conexões do AWS Direct Connect para obter mais informações.

Uma conexão Direct Connect usa interfaces virtuais (VIFs) para isolar diferentes fluxos de tráfego. Várias VIFs podem usar o mesmo link do Direct Connect, separadas por tags de VLAN (802.1q). Há três tipos de VIFs que fornecem conectividade à rede da AWS. Consulte Interfaces virtuais do AWS Direct Connect para obter mais detalhes. Os três tipos são:

- VIF privada: uma VIF privada é uma conexão privada entre seu dispositivo e seus recursos dentro da AWS. Elas terminam diretamente dentro da AWS em um Virtual Private Gateway (VGW) (que

oferece suporte a uma única VPC) ou por meio de um Direct Connect Gateway que se conecta a vários VGWs.

- VIF pública: uma VIF pública permite a conectividade com qualquer recurso público da AWS, como S3, DynamoDB e intervalos de IP públicos do EC2. Embora uma VIF pública não tenha acesso direto à Internet, qualquer recurso público da Amazon pode acessá-la (incluindo instâncias EC2 públicas de outros clientes), o que os clientes devem considerar durante o planejamento de segurança.
- VIF de trânsito: uma VIF de trânsito é uma conexão privada entre seu dispositivo e um AWS Transit Gateway por meio de um Direct Connect Gateway. Os VIFs de trânsito agora têm suporte em links com velocidades inferiores a 1 Gbps. Consulte o anúncio de lançamento para obter mais detalhes.

> **Note**
>
> A Hosted Virtual Interface (Hosted VIF) é um tipo de VIF privada em que a VIF é atribuída a uma conta da Conta da AWS diferente da Conta da AWS que possui a conexão do AWS Direct Connect (que pode incluir um parceiro do AWS Direct Connect). A AWS não permite mais que novos parceiros ofereçam esse modelo. Para obter mais informações, consulte Criar uma interface virtual hospedada.

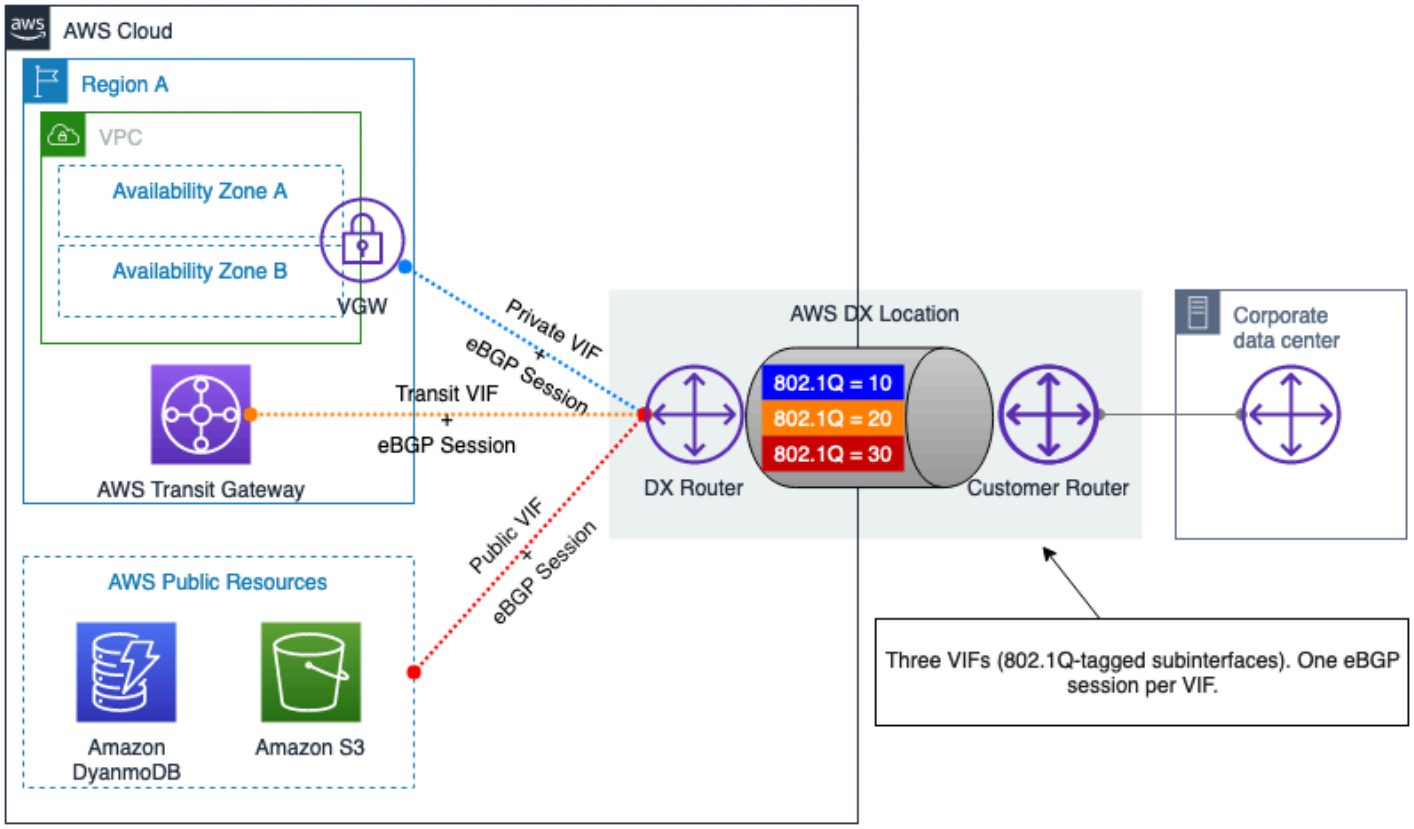

Figura 1 – VIFs privadas e públicas doAWS Direct Connect

Site-to-Site Virtual Private Network (VPN)

Uma VPN site a site permite que duas redes se comuniquem com segurança e pode ser usada em um transporte não confiável, como a Internet. Os clientes podem estabelecer conexões VPN entre sites on-premises e Amazon Virtual Private Clouds (Amazon VPC) por meio de duas opções:

- AWS Managed Site-to-Site VPN (AWS S2S VPN): este é um serviço de VPN totalmente gerenciado e altamente disponível que utiliza IPSec. Consulte O que é o AWS Site-to-Site VPN para obter mais informações. Opcionalmente, você pode habilitar a aceleração para sua conexão do Site-to-Site VPN. Consulte Conexões VPN site a site aceleradas para obter mais informações. A S2S VPN também pode utilizar VIFs de trânsito do Direct Connect para evitar que o tráfego percorra a Internet, reduzindo custos e permitindo o uso de endereços IP privados. Para obter detalhes, consulte VPN de IP privado com o AWS Direct Connect.
- Software Site-to-Site VPN (Customer-managed VPN): com essa opção de conectividade VPN, você é responsável pelo provisionamento e gerenciamento de toda a solução VPN, normalmente

executando software VPN em uma instância EC2. Para obter mais informações, consulte Software Site-to-Site VPN.

Ambas as opções exigem suporte no dispositivo de gateway do cliente para encerrar a extremidade on-premises dos túneis VPN. Esse dispositivo pode ser um dispositivo físico ou um software. ara obter mais informações sobre dispositivos de rede testados pela AWS, consulte a lista de dispositivos de gateway do cliente testados.

Transit Gateway Connect (TGW Connect)

O Transit Gateway Connect usa túneis GRE entre um AWS Transit Gateway e um dispositivo de gateway on-premises. O BGP é utilizado sobre o TGW Connect para permitir roteamento dinâmico. Observe que o TGW Connect não é criptografado. Para obter mais informações, consulte Transit Gateway Connect.

Serviços de conectividade híbrida da AWS

Os serviços de conectividade híbrida da AWS fornecem componentes de rede altamente escaláveis e altamente disponíveis. Eles desempenham um papel essencial na criação de soluções de rede híbrida. No momento da redação deste whitepaper, havia três endpoints de serviço principais:

- O AWS Virtual Private Gateway (VGW) é um serviço regional altamente redundante que fornece roteamento IP e encaminhamento no nível da VPC, atuando como o gateway para a VPC se comunicar com seus dispositivos de gateway do cliente. O VGW pode encerrar conexões do AWS S2S VPN e VIFs privados do AWS Direct Connect.
- O AWS Transit Gateway (TGW) é um serviço regional, altamente disponível e escalável, que permite a conexão entre várias VPCs, bem como suas redes locais por meio de VPN site a site e/ou Direct Connect, utilizando um único gateway centralizado. Conceitualmente, um AWS Transit Gateway atua como um roteador de nuvem virtual altamente disponível e redundante. O AWS Transit Gateway oferece suporte a roteamento de vários caminhos (ECMP) de custo igual em várias conexões Direct Connect, túneis VPN ou pares do TGW Connect. Os Transit Gateways podem se comunicar entre si, tanto na mesma região quanto entre regiões, permitindo que seus recursos conectados se comuniquem pelos links de peering. Para obter mais detalhes, consulte Cenários do AWS Transit Gateway.
- O Nuvem AWS WAN fornece um painel central para estabelecer conexões entre seus escritórios filiais, data centers e Amazon VPCs, construindo uma rede global com apenas alguns cliques.

Você usa políticas de rede para automatizar as tarefas de gerenciamento e segurança da rede em um único local. Para obter mais detalhes, consulte Documentação do Nuvem AWS WAN.

- O Direct Connect Gateway (DXGW) é um serviço disponível globalmente que distribui informações de roteamento em suas conexões, comportando-se de forma semelhante aos refletores de rota BGP em uma rede tradicional. Os dados não passam por um DXGW, ele apenas lida com informações de roteamento. Você pode criar um DXGW em qualquer Região da AWS e acessá-lo de todas as outras Regiões da AWS. Você pode conectar VIFs do Direct Connect a um DXGW e, em seguida, associar o DXGW a VGWs (usando VIFs privadas) ou a um AWS Transit Gateway (usando VIFs de trânsito). Para obter mais informações, consulte Gateways do Direct Connect. Você não precisa criar vários DXGWs para redundância, pois é um serviço de disponibilidade global. No entanto, você pode optar por usar vários DXGWs para separar os domínios de roteamento, por exemplo, uma rede de produção e uma rede de teste que você deseja manter completamente isoladas.

Tipo de conectividade híbrida e considerações sobre o design

Esta seção do whitepaper aborda as considerações que afetam suas escolhas ao selecionar uma rede híbrida à qual conectar seus ambientes on-premises à AWS. Ela segue um processo de pensamento lógico para ajudá-lo a selecionar uma solução de conectividade híbrida ideal. As considerações que afetam seu projeto são categorizadas em considerações que afetam seu tipo de conectividade e considerações que afetam seu design de conectividade. As considerações sobre o tipo de conectividade ajudarão você a decidir entre usar uma VPN baseada na Internet ou o Direct Connect. As considerações de design de conectividade ajudarão você a decidir como configurar as conexões.

As seguintes considerações que afetam seu tipo de conectividade são abordadas: tempo para implantação, segurança, SLA, desempenho e custo. Depois de analisar essas considerações e como elas afetam suas escolhas de design, você poderá decidir se é recomendável usar uma conexão baseada na Internet ou o Direct Connect para atender às suas necessidades.

As seguintes considerações que afetam seu design de conectividade são abordadas: escalabilidade, modelo de comunicação, confiabilidade e integração SD-WAN de terceiros. Depois de analisar essas considerações e como elas afetam suas escolhas de design, você poderá decidir o design lógico ideal recomendado para atender às suas necessidades.

A estrutura a seguir é usada para discutir e analisar cada uma das considerações de seleção e design:

- Definição - Breve definição do que é a consideração.
- Perguntas-chave - Fornece um conjunto de perguntas para permitir que você colete os requisitos associados à consideração.
- Capacidades a serem consideradas - Soluções para atender aos requisitos associados à consideração.
- Árvore de decisão - Para algumas considerações ou um grupo de considerações, uma árvore decisória é fornecida para ajudá-lo a selecionar a solução de rede híbrida ideal.

As considerações que afetam seu projeto de rede híbrida são abordadas em uma ordem em que a saída de uma consideração é parte da entrada para a consideração subsequente. Conforme

ilustrado na Figura 2, a primeira etapa é decidir sobre o tipo de conectividade, seguida de refiná-la com as considerações de seleção do projeto.

A Figura 2 demonstra as duas categorias de consideração, as considerações individuais e a ordem lógica na qual as considerações são abordadas nas subseções subsequentes. Essas são as considerações essenciais ao tomar uma decisão sobre o design da rede híbrida. Se o design direcionado não exigir todas essas considerações, você pode se concentrar nas considerações que se aplicam aos seus requisitos.

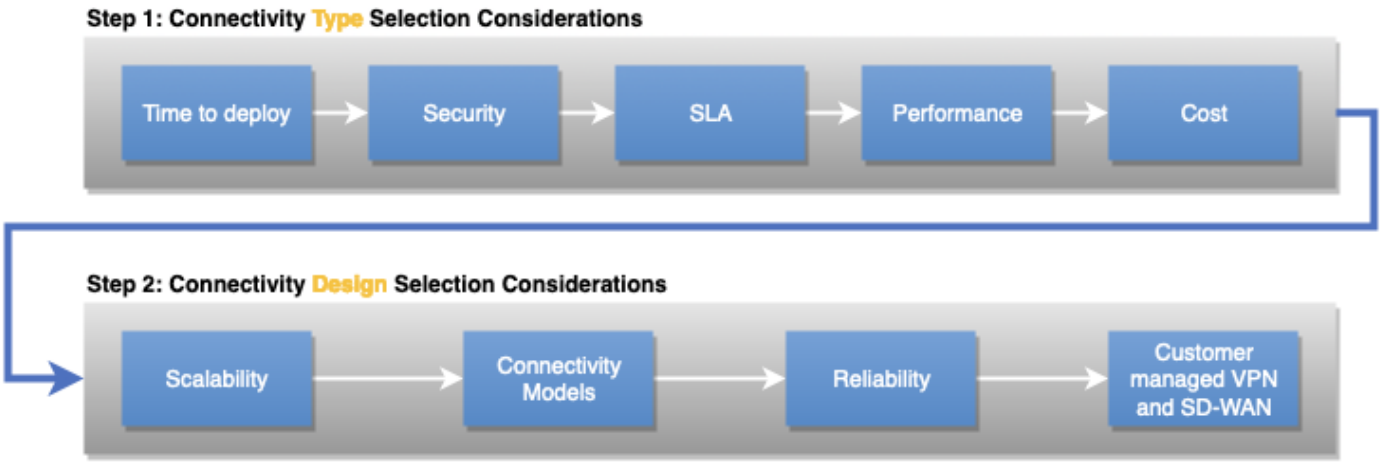

Figura 2 – Categorias de consideração, considerações individuais e a ordem lógica entre elas

Seleção do tipo de conectividade

Esta seção aborda considerações que afetam o tipo de conectividade que você seleciona para sua carga de trabalho. Isso inclui tempo de implantação, segurança, SLA, desempenho e custo.

Considerações

- Tempo para implantação
- Segurança
- Acordo de serviço (SLA)
- Desempenho
- Custo

Tempo para implantação

Definição

O tempo para implantação pode ser um fator importante na seleção de um tipo de conectividade adequado para uma carga de trabalho. Dependendo do tipo de conectividade e das localizações locais, a conectividade pode ser estabelecida em poucas horas; no entanto, pode levar semanas ou meses se circuitos adicionais precisarem ser instalados. Isso influenciará sua decisão de usar uma conexão baseada na Internet, uma conexão privada dedicada ou uma conexão hospedada privada fornecida como um serviço gerenciado por um parceiro do AWS Direct Connect.

Perguntas chave

- Qual é o cronograma necessário para a implantação: horas, dias, semanas ou meses?
- Por quanto tempo a conexão será necessária? Será um projeto de curta duração ou uma infraestrutura permanente?

Capacidades a serem consideradas

Quando você precisar de conectividade da AWS em horas ou dias, provavelmente precisará usar uma conexão de rede existente. Isso geralmente significa estabelecer uma conexão VPN pela Internet pública. AWS Se um parceiro AWS DX existente estiver fornecendo conectividade privada da AWS, uma nova conexão hospedada poderá ser provisionada em poucas horas.

Quando você tem dias ou semanas, pode trabalhar com um parceiro do AWS Direct Connect para estabelecer uma conectividade privada com a AWS. AWS Direct Connect Os parceiros ajudam você a estabelecer conectividade de rede entre os locais do AWS Direct Connect e seu datacenter, escritório ou ambiente de co-localização. Alguns Parceiros do AWS Direct Connect são aprovados para oferecer conexões hospedadas do Direct Connect. As conexões hospedadas geralmente podem ser provisionadas mais rapidamente do que as conexões dedicadas. AWS Direct Connect O parceiro provisionará cada conexão hospedada usando sua infraestrutura existente conectada ao backbone da AWS.

Quando você tiver várias semanas ou meses, poderá investigar o estabelecimento de uma conexão privada dedicada com a AWS. Provedores de serviços e parceiros do AWS Direct Connect facilitam conexões dedicadas do AWS Direct Connect. É comum que os provedores de serviços instalem equipamentos de rede nas instalações do cliente para facilitar uma conexão dedicada Direct

Connect. Dependendo do provedor de serviços, da localização do seu site e de outros fatores físicos, a instalação de uma conexão dedicada Direct Connect pode levar de várias semanas a alguns meses.

Se você já tem seus equipamentos de rede instalados no mesmo local de co-locação onde o local do AWS Direct Connect existe, é possível estabelecer rapidamente uma conexão dedicada do AWS Direct Connect por meio de uma interconexão no local de co-locação. Depois de solicitar a conexão, a AWS disponibiliza uma carta de autorização e atribuição de instalação de conexão (LOA-CFA) para download ou envia um e-mail solicitando mais informações. A LOA-CFA é a autorização para se conectar à AWS, sendo exigida pelo provedor de rede a fim de pedir uma conexão cruzada para você.

Tabela 1 – Comparação de efetividade de custos

	Conectividade baseada na Internet	Conexão dedicada DX (equipamento existente no local do DX)	Conexão dedicada DX (rede-nova)	Conexão hospedada DX (porta existente com o parceiro do DX)	Conexão hospedada DX (rede-nova)
Tempo de provisionamento	Horas a dias	Dias	Várias semanas a meses	Horas a dias	Vários dias a semanas ou meses

> **Note**
>
> As diretrizes de tempo de provisão fornecidas são baseadas na observação do mundo real e servem apenas como ilustração. Ao considerar a localização do site, a proximidade dos locais de conexão direta e a infraestrutura preexistente, tudo isso afetará o tempo de provisionamento. Seu parceiro do AWS Direct Connect o aconselhará sobre o tempo exato de provisionamento.

Segurança

Definição

Os requisitos de segurança influenciarão seu tipo de conectividade híbrida. Essas considerações incluem:

- Tipo de transporte: conexão à Internet ou rede privada
- Requisitos de criptografia

Perguntas chave

- Seus requisitos e políticas de segurança permitem o uso de conexões criptografadas pela Internet para conexão com a AWS, ou exigem o uso de conexões de rede privadas?
- Ao aproveitar as conexões de rede privada, a camada de rede precisa fornecer criptografia em trânsito?

Soluções técnicas

Seus requisitos e políticas de segurança podem permitir o uso da Internet ou exigir o uso de uma conexão de rede privada entre a AWS a rede da sua empresa. Eles também afetam a decisão se a rede deve fornecer criptografia em trânsito ou se a criptografia na camada do aplicativo é aceitável.

Se você pode aproveitar a Internet, o AWS Site-to-Site VPN pode ser usado para criar túneis criptografados entre sua rede e suas Amazon VPCs ou AWS Transit Gateways pela Internet. Estender sua solução SD-WAN para a AWS Internet também é uma opção se você estiver aproveitando uma conexão baseada na Internet. A seção VPN e SD-WAN gerenciadas pelo cliente, mais adiante neste whitepaper, aborda as considerações específicas da SD-WAN.

Se você precisar de uma conexão de rede privada entre a AWS a rede da sua empresa, a AWS recomenda o uso de conexões dedicadas do AWS Direct Connect ou conexões hospedadas. Se a criptografia em trânsito for necessária em uma conexão de rede privada, você deverá estabelecer uma VPN pelo Direct Connect (seja por VIF pública ou VIF de trânsito) ou considerar usar o MACsec em uma conexão dedicada de 10 Gbps ou 100 Gbps.

Tabela 2 – Exemplo de requisitos de conectividade para uma corporação automotiva

	Site-to-Site VPN	Direct Connect
Transporte	Internet	Conexão de rede privada
Criptografia em trânsito	Sim	Requer S2S VPN sobre DX, S2S VPN em um VIF de trânsito ou MACsec em uma conexão dedicada de 10 Gbps ou 100 Gbps

Acordo de serviço (SLA)

Definição

As organizações corporativas geralmente exigem que um provedor de serviços cumpra um SLA para cada serviço que a organização consome. A organização, por sua vez, cria seus próprios serviços e pode oferecer a seus próprios consumidores um SLA. O SLA é importante porque descreve como o serviço é fornecido e operado e geralmente inclui características mensuráveis específicas, como disponibilidade. Se o serviço violar o SLA definido, um provedor de serviços geralmente oferece uma compensação financeira especificada pelo contrato. Um SLA define o tipo de medida, o requisito e o período de medição. Como exemplo, consulte a definição da meta de tempo de atividade no SLA do AWS Direct Connect.

Perguntas-chave

- É necessário um SLA de conexão de conectividade híbrida com créditos de serviço?
- Toda a rede híbrida precisa cumprir uma meta de tempo de atividade?

Capacidades a serem consideradas

Tipo de conectividade: a conectividade com a Internet pode ser imprevisível. Embora a AWS tenha muito cuidado com vários links estabelecidos com um conjunto diversificado de ISPs, a administração da Internet está simplesmente fora do domínio administrativo da AWS ou de um único provedor. Há uma quantidade limitada de engenharia de rotas e influência de tráfego que um provedor de nuvem pode exercer quando o tráfego sai da fronteira da rede. Dito isso, há um SLA do AWS Site-to-Site VPN que fornece metas de disponibilidade para endpoints do AWS Site-to-Site VPN.

O AWS Direct Connect oferece um SLA formal com créditos de serviço calculados como uma porcentagem do total de encargos de hora da porta do AWS Direct Connect pagos por você pelas conexões aplicáveis que experimentaram indisponibilidade durante o ciclo de faturamento mensal no qual o SLA não foi cumprido. Esse é o transporte recomendado se for necessário um SLA. O AWS Direct Connect lista os requisitos mínimos de configuração específicos para cada destino de tempo de atividade, como número de locais, conexões e outros detalhes de configuração do AWS Direct Connect. A falha em atender aos requisitos significa que os créditos de serviço não podem ser oferecidos caso o serviço quebre os SLAs definidos.

É importante ressaltar que, mesmo que o serviço selecionado para fornecer conectividade híbrida esteja configurado para atender aos requisitos do SLA, o restante da rede pode não fornecer o mesmo nível de SLA. A responsabilidade da AWS termina no local do AWS Direct Connect na porta do AWS Direct Connect. Depois que a AWS transfere o tráfego para a rede da sua organização, ele não é mais responsabilidade da AWS. Se você usa um provedor de serviços entre a AWS e sua rede local, a conectividade está sujeita ao SLA entre você e o provedor de serviços, se aplicável. Lembre-se de que toda a rede híbrida é tão boa quanto a parte mais fraca dela ao projetar a conectividade híbrida.

Os parceiros do AWS Direct Connect oferecem conectividade do AWS Direct Connect. O parceiro pode oferecer um SLA com créditos de serviço com base em sua oferta de produto até o ponto de demarcação com a AWS. A opção deve ser avaliada e pesquisada mais detalhadamente diretamente com os parceiros da APN. A AWS publica uma lista de parceiros de entrega validados.

Design lógico: além do tipo de conectividade, você também deve considerar outros elementos básicos como parte de seu design geral. Por exemplo, o AWS Transit Gatewaytem seu próprio SLA, assim como o AWS S2S VPN. Você pode usar o AWS Transit Gateway para escalar e o AWS S2S VPN por razões de segurança, mas deve projetar ambos de maneira consistente com cada SLA para ser elegível para créditos de serviço com cada serviço respectivo.

Analise as recomendações de resiliência e o kit ferramentas de resiliência do AWS Direct Connect.

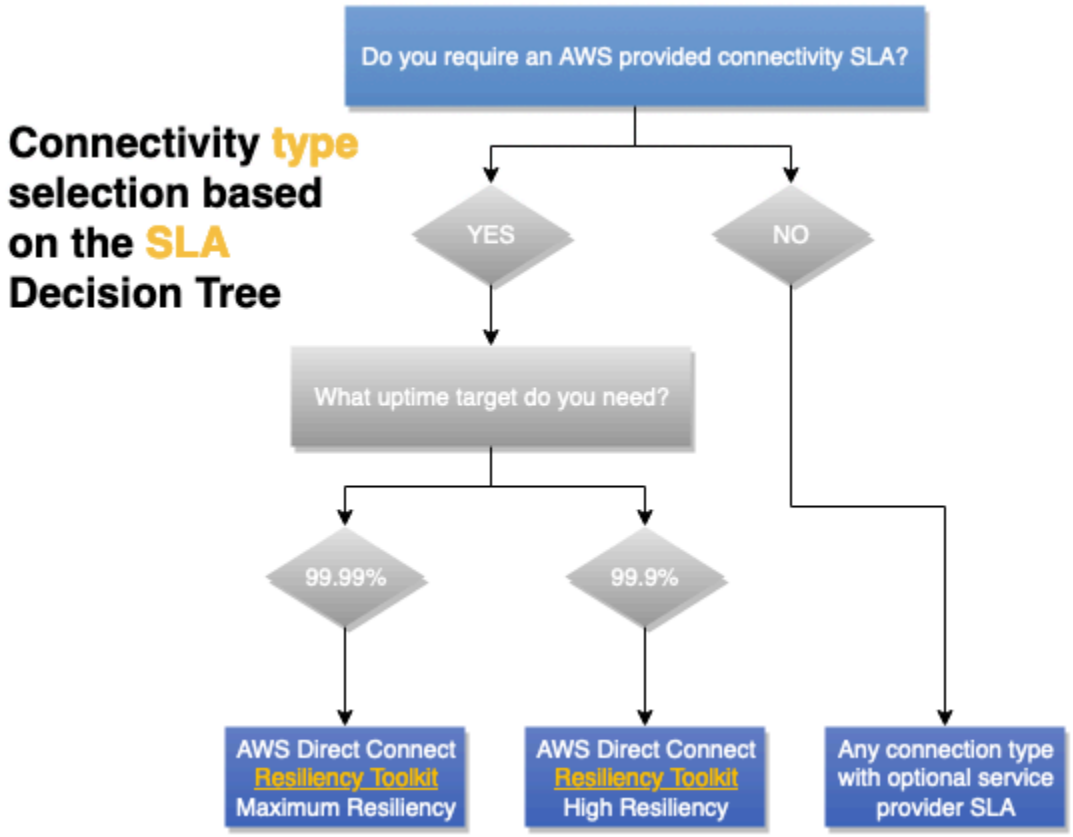

Figura 3 – Árvore de decisão de consideração do SLA

Desempenho

Definição

Há vários fatores que influenciam o desempenho da rede, como latência, perda de pacotes, instabilidade e largura de banda. Dependendo dos requisitos do aplicativo, a importância de cada um desses fatores pode variar.

Perguntas-chave

Com base nos requisitos do seu aplicativo, você precisa identificar e priorizar os fatores de desempenho da rede que afetam o comportamento do aplicativo e a experiência do usuário.

Largura de banda

A largura de banda se refere à taxa de transferência de dados de uma conexão e geralmente é medida em bits por segundo (bps). Megabits por segundo (Mbps) e gigabits por segundo (Gbps) são

escalas comuns e são de base 10 (1.000.000 de bits por segundo = 1 Mbps) em oposição à base 2 (2^10) vista em outros lugares.

Ao avaliar as necessidades de largura de banda dos aplicativos, lembre-se de que os requisitos de largura de banda podem mudar com o tempo. A implantação inicial na nuvem, as operações normais, as novas cargas de trabalho e os cenários de failover podem ter diferentes requisitos de largura de banda.

Os aplicativos podem ter suas próprias considerações sobre largura de banda. Alguns aplicativos podem exigir desempenho determinístico em uma conexão de alta largura de banda, enquanto outros podem exigir desempenho determinístico e alta largura de banda. Um aplicativo pode precisar de uma configuração especial para usar vários fluxos de tráfego (às vezes chamados de fluxos ou soquetes) em paralelo se estiver atingindo os limites de largura de banda por fluxo de tráfego, permitindo que use mais largura de banda da conexão. As VPNs podem limitar a taxa de transferência devido a sobrecargas de tunelamento, limites mais baixos de MTU ou limitações de largura de banda de hardware.

Latência

A latência é o tempo necessário para que um pacote vá da origem ao destino por meio de uma conexão de rede e geralmente é medida em milissegundos (ms), com requisitos de baixa latência às vezes expressos em microssegundos (µs). A latência é uma função da velocidade da luz, portanto, a latência aumenta com a distância.

Os requisitos de latência do aplicativo podem assumir diferentes formas. Um aplicativo altamente interativo, como um desktop virtual, pode ter uma meta de latência medida desde o momento em que o usuário executa uma entrada até que o usuário veja a área de trabalho virtual reagir a essa entrada. Os aplicativos de voz sobre IP (VoIP) podem ter requisitos semelhantes. Um segundo tipo de workload a ser considerada são aqueles que são altamente transacionais, que precisam de uma resposta do servidor antes de poderem continuar. Bancos de dados ou outras formas de armazenamento de chave/valor podem ser altamente afetados pelo aumento da latência da rede.

Jitter

O jitter mede a consistência da latência da rede e, assim como a latência, geralmente é medido em milissegundos (ms).

Os requisitos de jitter do aplicativo geralmente são encontrados em aplicativos de streaming em tempo real, incluindo entrega de vídeo e voz. Esses aplicativos tendem a exigir que seu fluxo de

dados esteja em uma taxa e atraso consistentes, com pequenos buffers para corrigir pequenas quantidades de instabilidade.

Perda de pacotes

A perda de pacotes é a medida de qual porcentagem do tráfego de rede não é entregue. Às vezes, todas as redes apresentam algum grau de perda de pacotes devido a altas explosões de tráfego, reduções de capacidade, falhas no equipamento de rede e outros motivos. Portanto, os aplicativos devem ter alguma tolerância à perda de pacotes; no entanto, o quanto eles podem tolerar pode variar de aplicativo para aplicativo.

Os aplicativos que usam TCP para transportar seu tráfego têm a capacidade de corrigir a perda de pacotes por meio de retransmissão. Aplicativos que utilizam UDP ou seus próprios protocolos sobre o IP precisam implementar seus próprios meios de lidar com a perda de pacotes e podem ser altamente sensíveis a ela. Um aplicativo de voz sobre IP pode simplesmente inserir silêncio na parte da chamada que teve o pacote perdido, em vez de tentar uma retransmissão. Algumas soluções de VPN incluem seus próprios mecanismos para se recuperar da perda de pacotes na rede que estão usando para transportar tráfego.

Capacidades a serem consideradas

Quando latência e throughput previsíveis são necessárias, o AWS Direct Connect é a escolha recomendada, pois fornece desempenho determinístico. A largura de banda pode ser selecionada com base nos requisitos de throughput. A AWS recomenda usar o AWS Direct Connect quando você precisar de uma experiência de rede mais consistente do que as conexões baseadas na Internet podem oferecer. Os VIFs privados e os VIFs de trânsito oferecem suporte a quadros gigantes, o que pode reduzir o número de pacotes na rede e melhorar a taxa de transferência devido à redução da sobrecarga. AWS Direct Connect O SiteLink permite usar o backbone da AWS para fornecer conectividade entre seus locais e pode ser habilitado sob demanda. A largura de banda usada para o SiteLink deve ser levada em consideração na seleção da largura de banda do Direct Connect.

Usar uma VPN em vez de AWS Direct Connect adiciona criptografia. No entanto, ela reduz o tamanho da MTU, o que pode reduzir a taxa de transferência. As capacidades do Site-to-Site VPN (S2S) gerenciada da AWS podem ser encontradas na documentação do AWS Site-to-Site VPN. Muitos locais de conexão direta oferecem suporte ao MACsec se a criptografia em sua conexão for o principal requisito de criptografia. O MACsec não tem as mesmas considerações de MTU ou de throughput potencial das conexões do Site-to-Site VPN. A AWS Transit Gateway permite que os clientes escalem horizontalmente o número de conexões VPN e aumentem a throughput de acordo com o Equal-cost Multipath Routing (ECMP). A VPN Site-to-Site gerenciada da AWS oferece suporte

ao uso de VIFs de trânsito do Direct Connect para conectividade privada. Consulte a VPN do IP privado com o AWS Direct Connect para obter detalhes.

Outra opção é usar uma VPN Site-to-Site gerenciada daAWS pela Internet. Pode ser uma opção atraente devido ao baixo custo e está amplamente disponível. No entanto, tenha em mente que o desempenho pela Internet é feito com o melhor esforço possível. Eventos climáticos na Internet, congestionamentos e períodos de maior latência podem ser imprevisíveis. A AWS oferece uma solução com o AWSAccelerated S2S VPN, que pode mitigar algumas das desvantagens de usar um caminho de internet. O Accelerated S2S VPN usa o AWS Global Accelerator, que permite que o tráfego de VPN entre na AWS rede o mais cedo possível e o mais próximo possível do dispositivo de gateway do cliente. Isso otimiza o caminho da rede, usando a rede global da AWS sem congestionamento para rotear o tráfego para o endpoint que fornece o melhor desempenho. Você pode utilizar conexões VPN aceleradas para evitar interrupções na rede que podem ocorrer quando o tráfego é roteado pela Internet pública.

Custo

Definição

Na nuvem, o custo da conectividade híbrida inclui o custo dos recursos provisionados e do uso. O custo dos recursos provisionados é medido em unidades de tempo, geralmente de hora em hora. O uso é para transferência e processamento de dados, geralmente medido em gigabytes (GB). Outros custos incluem o custo da conectividade com o ponto de presença da rede da AWS. Se sua rede estiver dentro da mesma instalação de colocalização, o custo pode ser tão baixo quanto o de uma interconexão. Se sua rede estiver em um local diferente, haverá custos envolvidos com um provedor de serviços ou parceiro do APN Direct Connect.

Perguntas-chave

- Quantos dados você espera enviar para a AWS por mês a partir de suas instalações e da Internet?
- Quantos dados você espera enviar da AWS por mês para as suas instalações e para a Internet?
- Com que frequência esses valores mudarão?
- O que muda em um cenário de falha?

Capacidades a serem consideradas

Se você deseja executar workloads com muita largura de banda na AWS, o AWS Direct Connect pode reduzir seus custos de rede da AWS de duas maneiras. Primeiro, ao transferir dados de e para

a AWS diretamente, você pode reduzir os custos de largura de banda pagos ao seu provedor de serviços de Internet. Em segundo lugar, todos os dados transferidos pela sua conexão dedicada são cobrados de acordo com a taxa de transferência de dados do AWS Direct Connect reduzida, em vez das taxas de transferência de dados da Internet. Consulte a página de preços do Direct Connect para obter detalhes.

O AWS Direct Connect permite o uso do AWS Direct Connect SiteLink para interconectar seus sites usando o backbone da AWS. Consulte o blog de lançamento do SiteLink para obter mais informações. O aproveitamento desse recurso gera custos normais de transferência de dados do Direct Connect, além de uma cobrança por hora em que o SiteLink está habilitado. Você pode habilitar e desabilitar o SiteLink sob demanda, e isso pode ser uma boa opção para cenários de falha envolvendo a conectividade com a Internet ou com a rede privada.

Se você estiver usando um provedor de serviços de rede para conectividade entre o local e um local do Direct Connect, sua capacidade e o tempo necessário para alterar seus compromissos de largura de banda se baseiam em seu contrato com o provedor de serviços.

O backbone da AWS pode entregar seu tráfego para qualquer Região da AWS, exceto a China, a partir de qualquer ponto de presença da rede da AWS. Esse recurso tem muitos benefícios técnicos em relação ao uso da Internet para acesso remoto a Regiões da AWS, mas tem um custo. Consulte a página de preços do EC2 Data Transfer para obter detalhes. Se houver um AWS Transit Gateway no caminho do tráfego, ele adicionará o custo de processamento de dados por GB. No entanto, se usar o emparelhamento entre regiões entre dois gateways de trânsito, você será cobrado apenas uma vez pelo processamento de dados do gateway de trânsito.

O design ideal do aplicativo mantém o processamento de dados dentro da AWS e minimiza as cobranças desnecessárias de saída de dados. A entrada de dados na AWS é gratuita.

> **Note**
>
> Como parte da solução geral de conectividade, além do custo de conexão da AWS, você também deve considerar o custo da conectividade de ponta a ponta, incluindo custo do provedor de serviços, conexões cruzadas, racks e equipamentos dentro do local DX (se necessário).

Se você não tiver certeza se deve usar a Internet ou uma conexão privada, calcule um ponto de equilíbrio no qual o AWS Direct Connect se torne mais barata do que usar a Internet. Se o volume

de dados significa que o AWS Direct Connect é mais barato e você precisa de conectividade permanente, o AWS Direct Connect é a melhor opção de conectividade.

Se a conectividade for temporária e a Internet atender a outros requisitos, pode ser mais barato usar o AWS S2S VPN pela Internet devido à elasticidade da Internet. Observe que isso requer que você tenha conectividade com a Internet suficiente de sua rede on-premises.

Se você estiver em uma instalação que tenha o AWS Direct Connect (a lista está disponível no site do Direct Connect), é possível estabelecer uma conexão cruzada com a AWS. Isso significa usar conexões dedicadas a 1,10 ou 100 Gbps. Os parceiros do AWS Direct Connect oferecem mais opções de largura de banda e capacidades menores, o que pode otimizar seu custo de conectividade. Por exemplo, você pode começar com uma conexão hospedada de 50 Mbps versus uma conexão dedicada de 1 Gbps.

Com o AWS Transit Gateway, você pode compartilhar suas conexões VPN e Direct Connect com várias VPCs. Embora você seja cobrado pelo número de conexões que faz com o AWS Transit Gateway por hora e pela quantidade de tráfego que flui por meio do AWS Transit Gateway, isso simplifica o gerenciamento e reduz o número de conexões VPN e VIFs necessárias. Os benefícios e a economia de custos da redução da sobrecarga operacional podem facilmente superar o custo adicional do processamento de dados. Opcionalmente, você pode considerar um design em que o AWS Transit Gateway esteja no caminho do tráfego para a maioria das VPCs, mas não para todas. Essa abordagem evita as taxas de processamento de dados do AWS Transit Gateway para casos de uso em que você precisa transferir grandes quantidades de dados para a AWS. Consulte a seção Modelos de conectividade para obter mais detalhes sobre esse design. Outra abordagem é combinar o AWS Direct Connect como caminho principal com o AWS S2S VPN pela Internet como caminho de backup/failover. Embora tecnicamente viável e muito econômica, essa solução tem desvantagens técnicas (discutidas na seção Confiabilidade deste whitepaper) e pode ser mais difícil de gerenciar. A AWS não recomenda isso para workloads altamente críticas ou críticas.

A abordagem final é uma VPN ou SD-WAN gerenciada pelo cliente implantada na(s) instância(s) do Amazon EC2. Isso pode ser mais barato em grande escala se houver dezenas a centenas de sites em comparação com o AWS S2S VPN. No entanto, há despesas gerais de gerenciamento, custos de licenciamento e custo de recursos do EC2 que cada dispositivo virtual deve ser considerado.

Matriz de decisão

Tabela 3 – Exemplo de entradas de design de conectividade para uma corporação automotiva

Categoria	VPN ou SD-WAN gerenciada pelo cliente	AWS S2S VPN	AWS S2S VPN acelerada	Conexão hospedada do AWS Direct Connect	Conexão dedicada do AWS Direct Connect
Requer conexão com a Internet	Sim	Sim	Sim	Não	Não
Custo dos recursos provisionados	Licenciamento de instância e software do EC2	AWS S2S VPN	AWS S2S VPN e AWS Global Accelerator	Fatia de capacidade aplicável do custo da porta	Custo da porta dedicada
Custo de transferência de dados	Taxa de Internet	Taxa de Internet ou taxa do Direct Connect	Internet com transferência de dados premium	Taxa do Direct Connect	Taxa do Direct Connect
Gateway de trânsito	Opcional	Opcional	Obrigatório	Opcional	Opcional
Custo de processamento de dados da AWS	N/D	Somente com AWS Transit Gateway	Sim	Somente com AWS Transit Gateway	Somente com AWS Transit Gateway
Pode ser usado sobre o AWS Direct Connect?	Sim	Sim	Não	N/D	N/D

Seleção de design de conectividade

Esta seção do whitepaper aborda as considerações que afetam a seleção do design de conectividade. O design de conectividade inclui os aspectos lógicos, bem como a forma de projetar e otimizar a confiabilidade da conectividade híbrida.

As seguintes considerações serão abordadas: escalabilidade, modelos de conectividade, confiabilidade e VPN e SD-WAN gerenciadas pelo cliente.

Considerações

- Escalabilidade
- Modelos de conectividade
- Confiabilidade
- VPN e SD-WAN gerenciadas pelo cliente

Escalabilidade

Definição

A escalabilidade se refere à capacidade de sua solução de conectividade crescer e evoluir ao longo do tempo à medida que seus requisitos mudam.

Ao projetar uma solução, você precisa considerar o tamanho atual, bem como o crescimento previsto. Esse crescimento pode ser orgânico ou estar relacionado à rápida expansão, como em cenários do tipo fusão e aquisição.

Observação: dependendo da arquitetura da solução desejada, talvez nem todos os elementos anteriores precisem ser levados em consideração. No entanto, eles podem servir como elementos fundamentais para identificar os requisitos de escalabilidade das soluções de rede híbrida mais comuns. Este whitepaper se concentra na seleção e no design da conectividade híbrida. É recomendável que você também considere a escala da conectividade híbrida em relação à arquitetura de rede VPC. Para obter mais informações, consulte o whitepaper Criando uma infraestrutura de AWS rede multi-VPC escalável e segura.

Perguntas chave

- Qual é o número atual e previsto de VPCs que exigem conectividade com sites on-premises?
- As VPCs são implantadas em uma Região da AWS ou várias regiões?

- Quantos sites on-premises precisam estar conectados à AWS?
- Quantos dispositivos de gateway do cliente (normalmente roteadores ou firewalls) você tem por site que precisam ser conectados à AWS?
- Quantas rotas devem ser anunciadas para as Amazon VPCs e qual é o número esperado de rotas que serão recebidas paralelamente? AWS
- É necessário aumentar a largura de banda ao AWS longo do tempo?

Capacidades a serem consideradas

A escala é um fator importante no design de conectividade híbrida. Até esse ponto, a seção subsequente incorporará a escala como parte do design do modelo de conectividade direcionado.

A seguir estão as melhores práticas recomendadas para minimizar a complexidade de escala do projeto de conectividade de rede híbrida:

- O resumo de rotas deve ser usado para reduzir o número de rotas anunciadas e recebidas. AWS Portanto, o esquema de endereçamento IP precisa ser projetado para maximizar o uso do resumo de rotas. A engenharia de tráfego é uma consideração geral fundamental. Para obter mais informações sobre engenharia de tráfego, consulte a subseção Engenharia de tráfego na seção Confiabilidade.
- Minimize o número de sessões de emparelhamento de BGP usando DXGW com VGW ou AWS Transit Gateway, onde uma única sessão de BGP pode fornecer conectividade a várias VPCs.
- Considere o Cloud WAN quando vários sites Regiões da AWS e locais precisarem estar conectados entre si.

Modelos de conectividade

Definição

O modelo de conectividade se refere ao padrão de comunicação entre as redes on-premises e os recursos de nuvem na AWS. Você pode implantar recursos de nuvem em uma Amazon VPC em uma Região da AWS ou várias VPCs em várias regiões, bem como AWS serviços que tenham um endpoint público em uma ou várias Regiões da AWS, como Amazon S3 e DynamoDB.

Perguntas chave

- Há um requisito para a comunicação entre VPCs dentro de uma região e entre regiões?

- Há algum requisito para acessar endpoints AWS públicos diretamente do local?
- Há um requisito para acessar AWS serviços usando VPC endpoints no local?

Capacidades a serem consideradas

Os seguintes são alguns dos cenários de modelos de conectividade mais comuns. Cada modelo de conectividade abrange requisitos, atributos e considerações.

Observação: conforme destacado anteriormente, este whitepaper se concentra na conectividade híbrida entre redes locais e a AWS. Para obter mais detalhes sobre o design para interconectar VPCs, consulte o whitepaper Construindo uma infraestrutura de rede AWS multi-VPC escalável e segura.

Modelos

- AWS VPN acelerada de site a site —, única AWS Transit GatewayRegião da AWS
- AWS DX — DXGW com VGW, região única
- AWS DX — DXGW com VGW, multirregiões e peering público AWS
- AWS DX — DXGW com AWS Transit Gateway, multirregiões e peering público AWS
- AWS DX — DXGW com AWS Transit Gateway, multirregiões (mais de 3)

AWS VPN acelerada de site a site —, única AWS Transit GatewayRegião da AWS

Este modelo é construído de:

- Solteiro Região da AWS.
- AWS Conexão VPN gerenciada de site a site com. AWS Transit Gateway
- VPN acelerada habilitada.

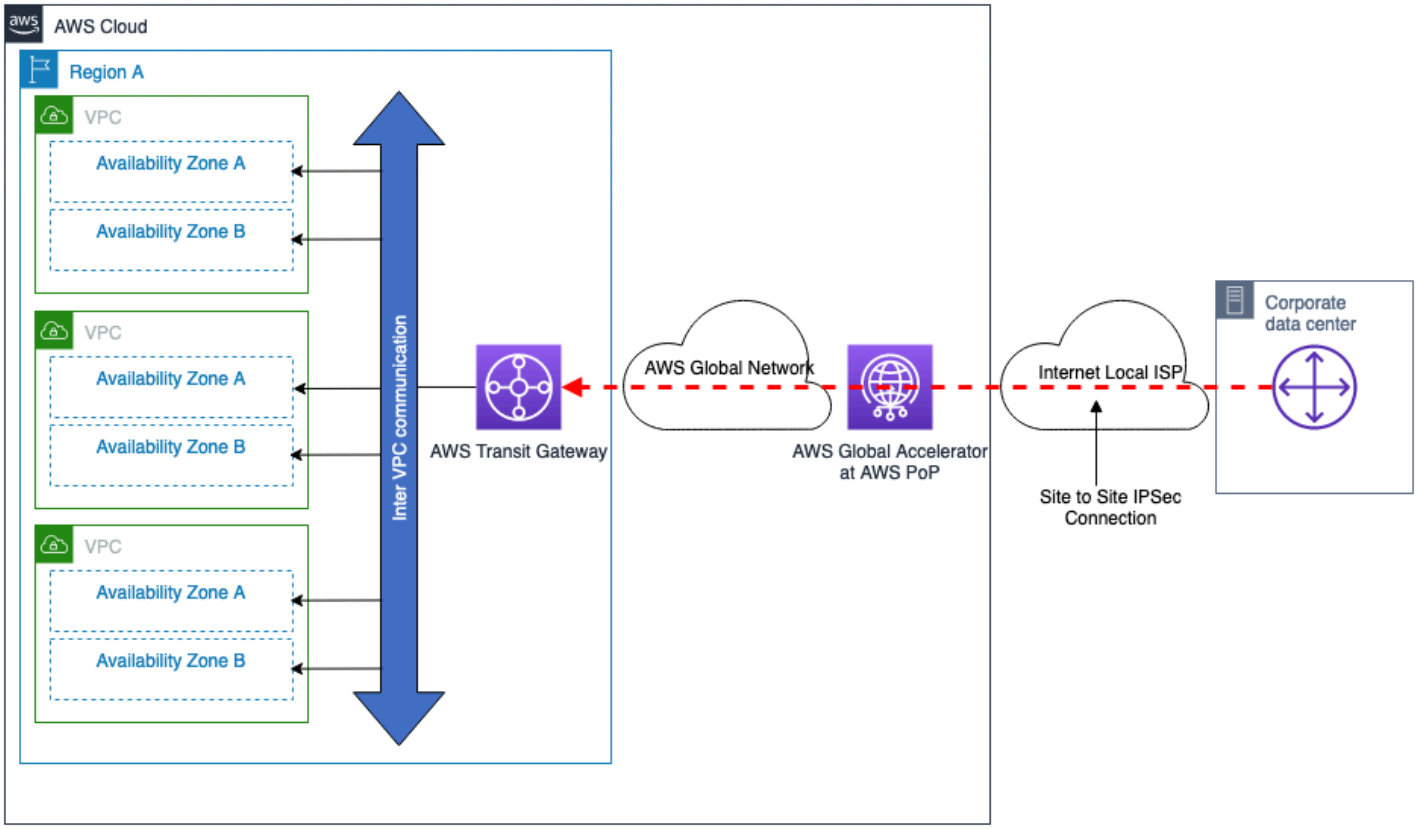

Figura 4 — VPN AWS gerenciada — AWS Transit Gateway, única Região da AWS

Atributos do modelo de conectividade:

- Forneça a capacidade de estabelecer conexões VPN otimizadas na Internet pública usando conexões VPN AWS site a site aceleradas.
- Forneça a capacidade de obter maior largura de banda de conexão VPN configurando vários túneis VPN com ECMP.
- Pode ser usado para conexão a partir de vários locais remotos.
- Oferece failover automático com roteamento dinâmico (BGP).
- Com a AWS Transit Gateway conexão às VPCs, todas as VPCs conectadas podem usar as mesmas conexões VPN. Você também pode controlar o modelo de comunicação desejado entre as VPCs. Para obter mais informações, consulte Como funcionam os gateways de trânsito.
- Oferece opções flexíveis de design para integrar dispositivos virtuais de segurança e SD-WAN de terceiros. AWS Transit Gateway Consulte Segurança de rede centralizada para tráfego de VPC para VPC e on-premises para VPC.

Considerações sobre escala:

- Até 50 Gbps de largura de banda com vários túneis IPsec e ECMP configurados (cada fluxo de tráfego será limitado à largura de banda máxima por túnel VPN).
- Milhares de VPCs podem ser conectadas por. AWS Transit Gateway
- Consulte as cotas VPN site a site para ver outros limites de escala, como o número de rotas.

Outras considerações:

- Os custos adicionais AWS Transit Gateway de processamento para transferência de dados entre o data center local e. AWS
- Grupos de segurança de uma VPC remota não podem ser referenciados em AWS Transit Gateway — no entanto, isso é suportado pelo emparelhamento de VPC.

AWS DX — DXGW com VGW, região única

Este modelo é construído de:

- Solteiro Região da AWS.
- AWS Direct Connect Conexões duplas para locais DX independentes.
- AWS DXGW conectado diretamente às VPCs usando VGW.
- Uso opcional de AWS Transit Gateway para comunicação entre VPC.

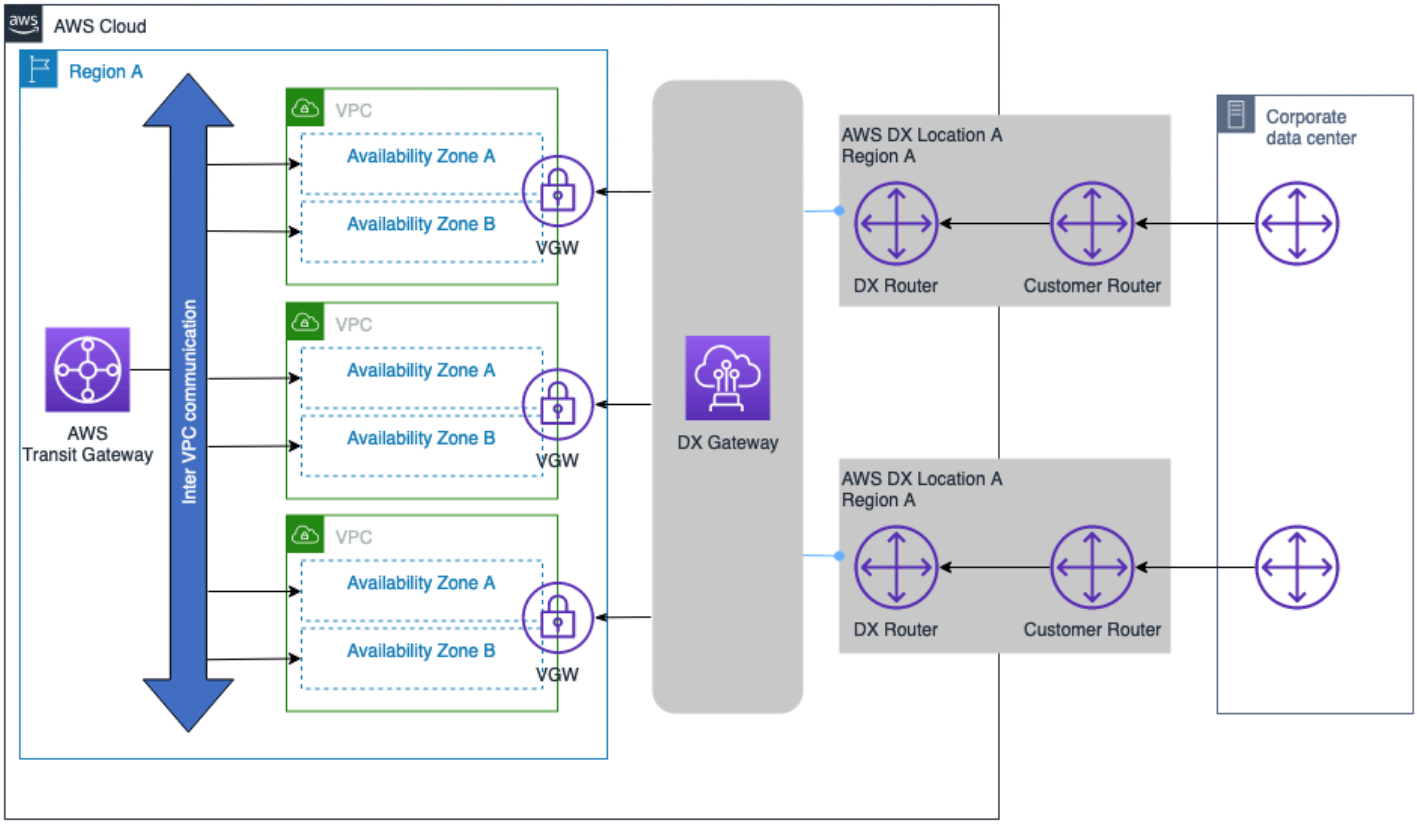

Figura 5 — AWS DX — DXGW com VGW, único Região da AWS

Atributos do modelo de conectividade:

- Fornece a capacidade de se conectar a VPCs e conexões DX em outras regiões no futuro.
- Oferece failover automático com roteamento dinâmico (BGP).
- Com AWS Transit Gateway você pode controlar o modelo de comunicação desejado entre as VPCs. Para obter mais informações, consulte Como funcionam os gateways de trânsito.

Considerações sobre escala:

Consulte as cotas do AWS Direct Connect para obter mais informações sobre outros limites de escala, como o número de prefixos compatíveis e o número de VIFs por tipo de conexão DX (dedicada, hospedada). Algumas considerações importantes:

- A sessão BGP para uma VIF privada pode anunciar até 100 rotas cada para IPv4 e IPv6.

- Até 20 VPCs podem ser conectadas por DXGW em uma única sessão de BGP. Se forem necessárias mais de 20 VPCs, é possível adicionar DXGWs adicionais para facilitar a conectividade em grande escala ou considerar o uso da integração do gateway de trânsito.
- Outros AWS Direct Connect s podem ser adicionados conforme desejado.

Outras considerações:

- Não incorre em custos de processamento AWS Transit Gateway relacionados à transferência de dados entre redes locais AWS e redes locais.
- Os grupos de segurança de uma VPC remota não podem ser referenciados (é AWS Transit Gateway necessário emparelhamento de VPC).
- O emparelhamento de VPC pode ser usado em vez de facilitar AWS Transit Gateway a comunicação entre as VPCs. No entanto, isso aumenta a complexidade operacional para criar e gerenciar um grande número de emparelhamentos de VPC em grande escala. point-to-point
- Se a comunicação entre VPC não for necessária, nem o emparelhamento de AWS Transit Gateway VPC será necessário nesse modelo de conectividade.

AWS DX — DXGW com VGW, multirregiões e peering público AWS

Este modelo é composto por:

- Vários data centers locais com conexões duplas a. AWS
- AWS Direct Connect Conexões duplas para locais DX independentes.
- AWS DXGW conectado diretamente a mais de 10 VPCs usando VGW, até 20 VPCs usando VGW.
- Uso opcional do AWS Transit Gateway para comunicação entre VPC e entre regiões.

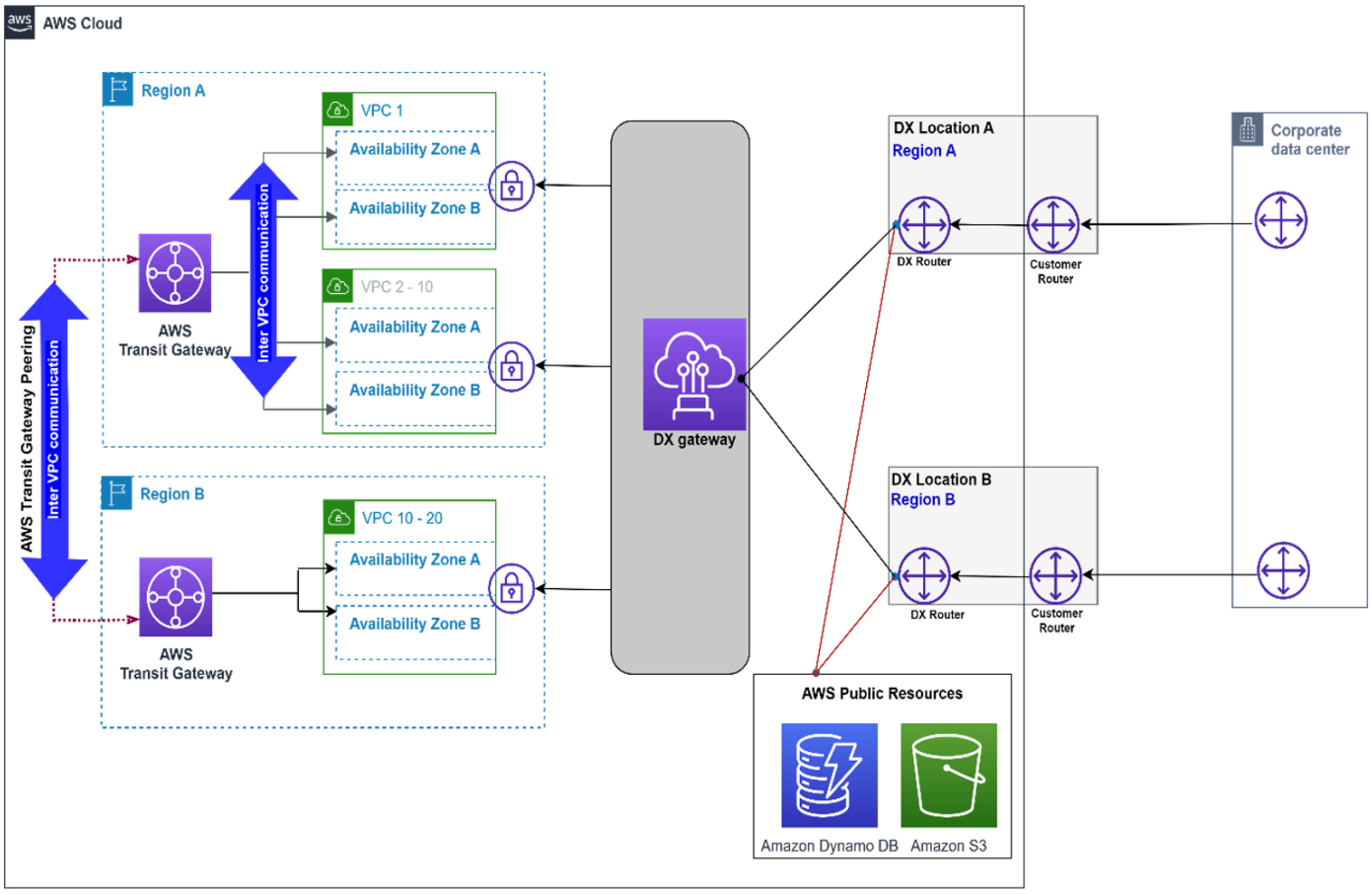

Figura 6 — AWS DX — DXGW com VGW, multirregiões e VIF pública

Atributos do modelo de conectividade:

- AWS DXGW conectado diretamente a mais de 10 VPCs usando VGW até 20 VPCs usando VGW.
- AWS O DX public VIF é usado para acessar serviços AWS públicos, como o Amazon S3, diretamente pelas conexões DX. AWS
- Fornece a capacidade de se conectar a VPCs e conexões DX em outras regiões no futuro.
- Comunicação VPC entre VPC e VPC entre regiões facilitada pelo AWS Transit Gateway peering do Transit Gateway.

Considerações sobre escala:

Consulte as cotas do AWS Direct Connect para obter mais informações sobre outros limites de escala, como o número de prefixos compatíveis e o número de VIFs por tipo de conexão DX (dedicada, hospedada). Algumas considerações importantes:

- A sessão BGP para uma VIF privada pode anunciar até 100 rotas cada para IPv4 e IPv6.
- Até 20 VPCs podem ser conectadas por DXGW em uma única sessão BGP em cada VIF privada, até 30 VIFs privadas por DXGW.
- Outros AWS Direct Connect s podem ser adicionados conforme desejado.

Outras considerações:

- Não incorre em custos de processamento AWS Transit Gateway relacionados à transferência de dados entre redes locais AWS e redes locais.
- Os grupos de segurança de uma VPC remota não podem ser referenciados por (é AWS Transit Gateway necessário emparelhamento de VPC).
- O emparelhamento de VPC pode ser usado em vez de facilitar AWS Transit Gateway a comunicação entre as VPCs. No entanto, isso aumentará a complexidade operacional para criar e gerenciar um grande número de emparelhamentos de VPC em grande escala. point-to-point
- Se a comunicação entre VPC não for necessária, nem o emparelhamento de AWS Transit Gateway VPC será necessário nesse modelo de conectividade.

AWS DX — DXGW com AWS Transit Gateway, multirregiões e peering público AWS

Este modelo é composto por:

- Múltiplo Regiões da AWS.
- AWS Direct Connect Conexões duplas para locais DX independentes.
- Um único data center local com conexões duplas com o. AWS
- AWS DXGW com. AWS Transit Gateway
- Alta escala de VPCs por região.

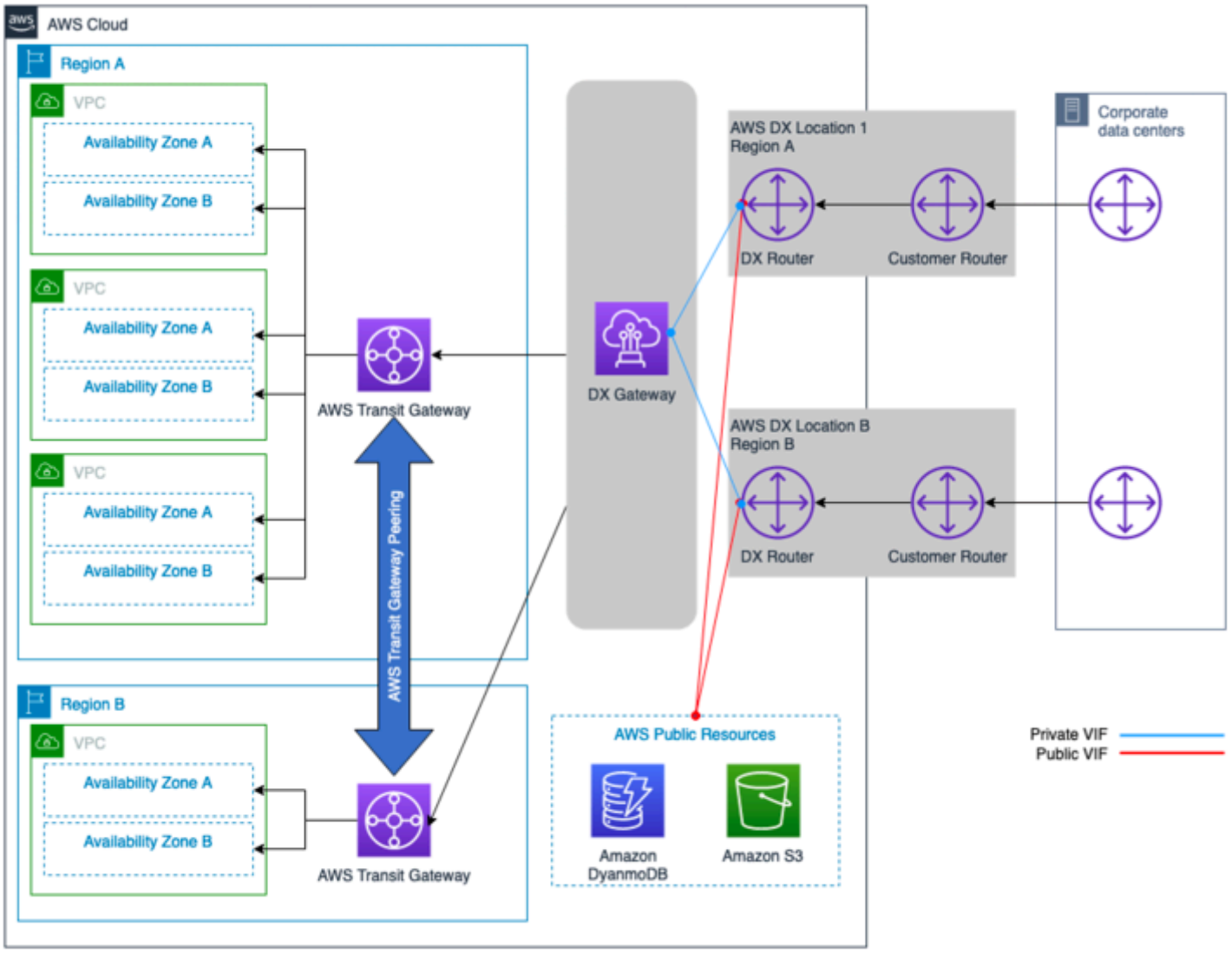

Figura 7 — AWS DX — DXGW com AWS Transit Gateway, multirregiões e VIF pública AWS

Atributos do modelo de conectividade:

- AWS O DX public VIF é usado para acessar recursos AWS públicos, como o S3, diretamente pelas conexões DX. AWS
- Fornece a capacidade de se conectar a VPCs e/ou conexões DX em outras regiões no futuro.
- Com a AWS Transit Gateway conexão às VPCs, a conectividade de malha total ou parcial pode ser alcançada entre as VPCs.
- Comunicação entre VPC e VPC entre regiões facilitada pelo peering. AWS Transit Gateway

- Oferece opções flexíveis de design para integrar dispositivos virtuais SDWAN e segurança de terceiros com AWS Transit Gateway. Consulte: Segurança de rede centralizada para tráfego de VPC para VPC e on-premises para VPC.

Considerações sobre escala:

- O número de rotas de ida e volta AWS Transit Gateway é limitado ao número máximo suportado de rotas em uma VIF de trânsito (os números de entrada e saída variam). Consulte as cotas do AWS Direct Connect para obter mais informações sobre os limites de escala e o número compatível de rotas e VIFs.
- Expanda até milhares de VPCs por AWS Transit Gateway mais de uma única sessão de BGP.
- VIF de trânsito único por AWS DX.
- Conexões AWS DX adicionais podem ser adicionadas conforme desejado.

Outras considerações:

- Incorre em custos adicionais AWS Transit Gateway de processamento para transferência de dados entre um AWS site local.
- Os grupos de segurança de uma VPC remota não podem ser referenciados por (é AWS Transit Gateway necessário emparelhamento de VPC).
- O emparelhamento de VPC pode ser usado em vez de facilitar AWS Transit Gateway a comunicação entre as VPCs. No entanto, isso aumentará a complexidade operacional para criar e gerenciar um grande número de emparelhamentos de VPC em grande escala. point-to-point
- Se forem necessários mais de três AWS Transit Gateway s, é possível adicionar DXGW adicional — consulte o seguinte modo de conectividade.

AWS DX — DXGW com AWS Transit Gateway, multirregiões (mais de 3)

Este modelo é construído de:

- Múltiplo Regiões da AWS (mais de 3).
- Datacenters on-premises duplos.
- AWS Direct Connect Conexões duplas com locais DX independentes por região.
- AWS DXGW com. AWS Transit Gateway
- Alta escala de VPCs por região.

- Malha completa de observação entre AWS Transit Gateway s.

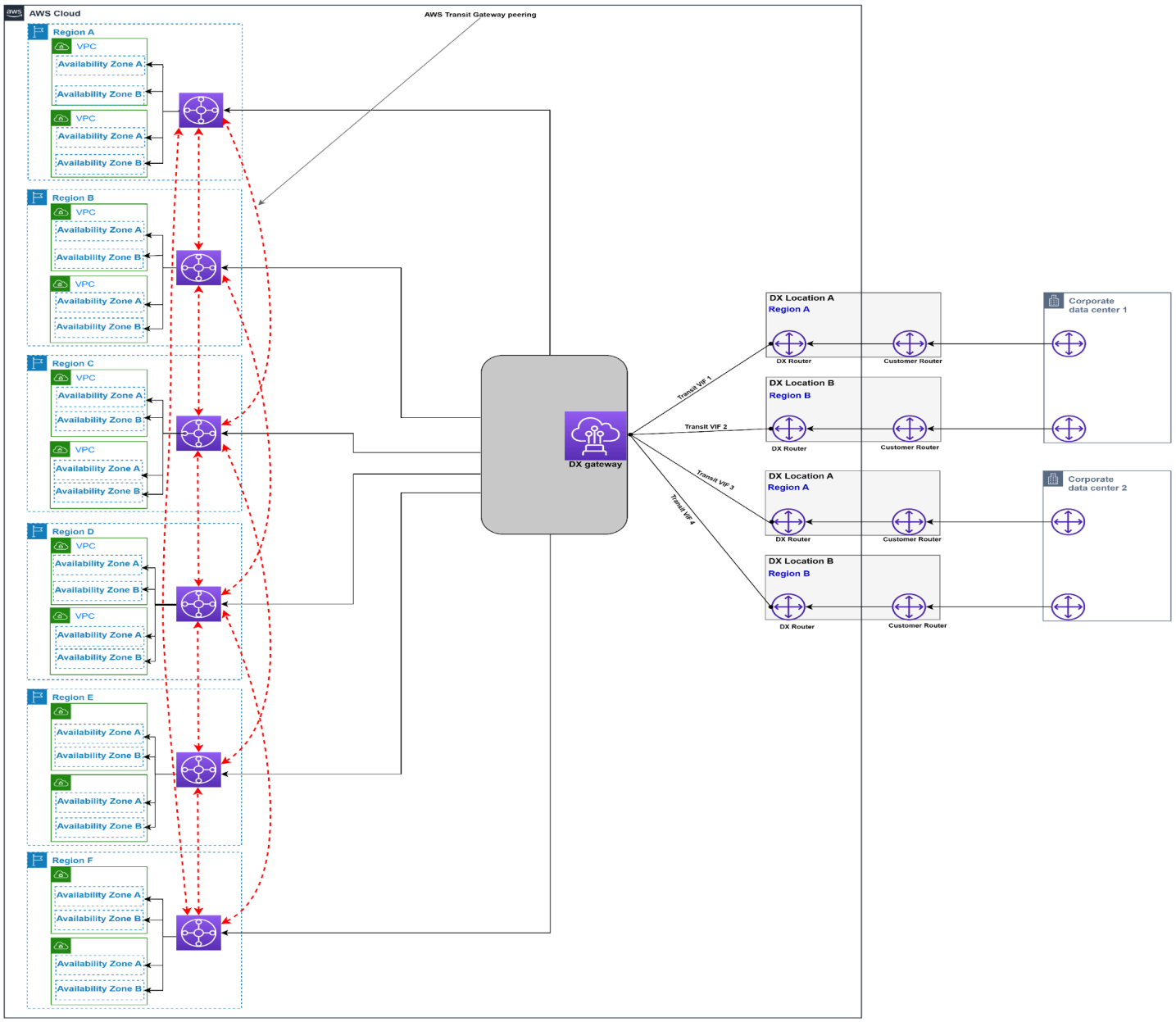

Figura 8 — AWS DX — DXGW com AWS Transit Gateway, multirregiões (mais de três)

Atributos do modelo de conectividade:

- Menor sobrecarga operacional.
- AWS O DX public VIF é usado para acessar recursos AWS públicos, como o S3, diretamente pelas conexões DX. AWS

- Fornece a capacidade de se conectar a VPCs e conexões DX em outras regiões no futuro.
- Com a AWS Transit Gateway conexão às VPCs, a conectividade de malha total ou parcial pode ser alcançada entre as VPCs.
- A comunicação VPC entre regiões é AWS Transit Gateway facilitada pelo peering.
- Oferece opções flexíveis de design para integrar dispositivos virtuais SDWAN e segurança de terceiros com AWS Transit Gateway. Consulte: Segurança de rede centralizada para tráfego de VPC para VPC e on-premises para VPC.

Considerações sobre escala:

- O número de rotas de ida e volta AWS Transit Gateway é limitado ao número máximo suportado de rotas em uma VIF de trânsito (os números de entrada e saída variam). Consulte as cotas do AWS Direct Connect para obter mais informações sobre os limites de escala. Considere o resumo da rota, se necessário, para reduzir o número de rotas.
- Expanda até milhares de VPCs por AWS Transit Gateway mais de uma única sessão de BGP por DXGW (supondo que o desempenho fornecido pelas conexões DX AWS provisionadas seja suficiente).
- Até seis AWS Transit Gateway s podem ser conectados por DXGW.
- Se mais de três regiões precisarem ser conectadas usando AWS Transit Gateway, serão necessários DXGWs adicionais.
- VIF de trânsito único por AWS DX.
- Conexões AWS DX adicionais podem ser adicionadas conforme desejado.

Outras considerações:

- Incorre em custos adicionais AWS Transit Gateway de processamento para transferência de dados entre o site local e. AWS
- Os grupos de segurança de uma VPC remota não podem ser referenciados por (é AWS Transit Gateway necessário emparelhamento de VPC).
- O emparelhamento de VPC pode ser usado em vez de facilitar AWS Transit Gateway a comunicação entre as VPCs. No entanto, isso aumentará a complexidade operacional para criar e gerenciar um grande número de emparelhamentos de VPC em grande escala. point-to-point

A árvore decisória a seguir abrange as considerações do modelo de escalabilidade e comunicação:

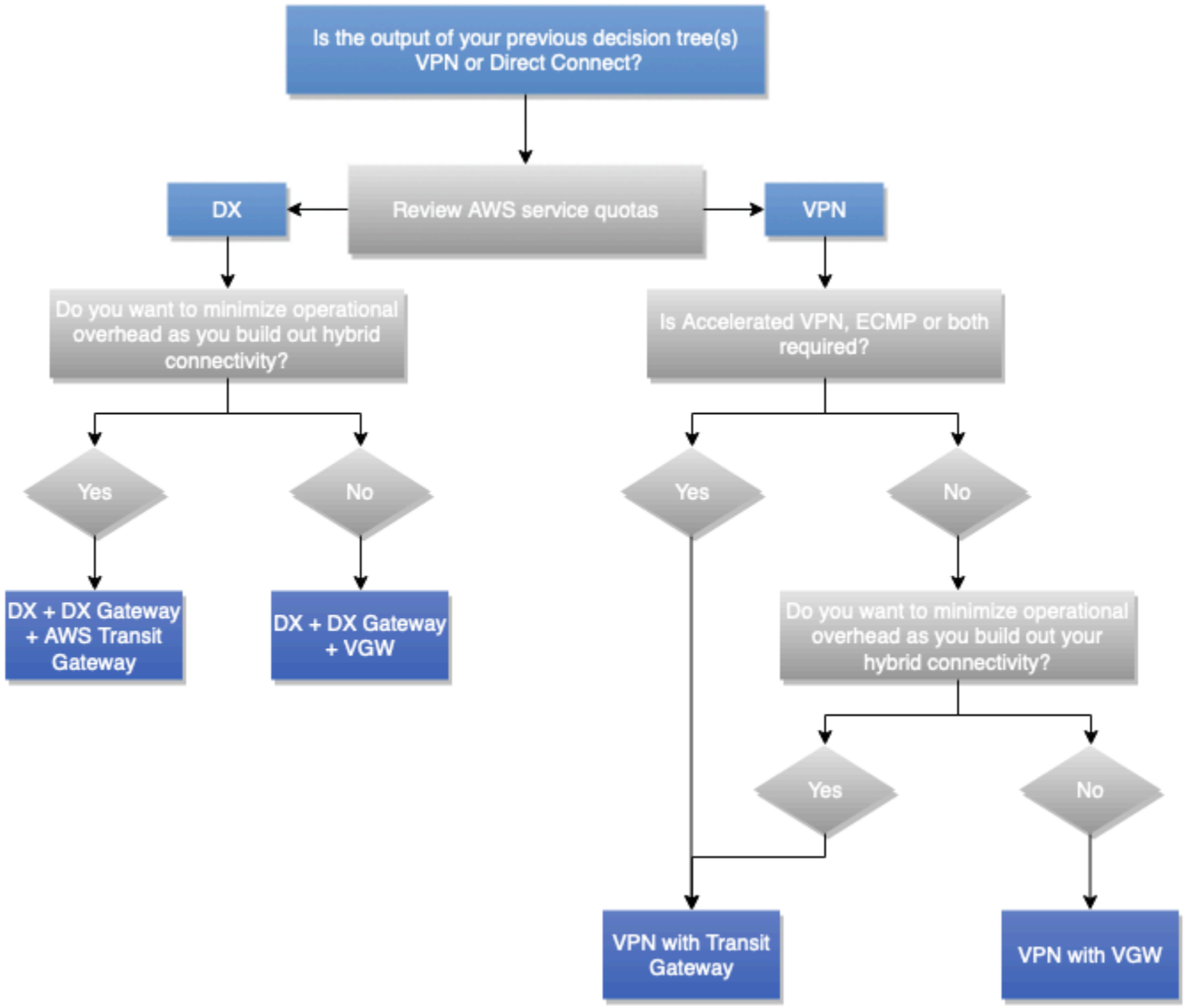

Figura 9 – Árvore de decisão para escalabilidade e modelo de comunicação

> **Note**
>
> Se o tipo de conexão selecionado for VPN, normalmente considerando o desempenho, deve-se decidir se o ponto de terminação da VPN é uma conexão VPN AWS VGW ou AWS Transit Gateway AWS S2S. Se ainda não foi decidido, então você pode considerar o modelo de comunicação necessário entre as VPCs, juntamente com o número de VPCs necessárias a serem conectadas à(s) conexão(ões) VPN, para ajudá-lo a tomar a decisão.

Confiabilidade

Definição

Confiabilidade se refere à capacidade de um serviço ou sistema de realizar sua função esperada quando necessário. A confiabilidade de um sistema pode ser medida pelo nível de sua qualidade operacional dentro de um determinado período de tempo. Compare isso com a resiliência, que se refere à capacidade de um sistema se recuperar de interrupções na infraestrutura ou no serviço de forma dinâmica e confiável.

Para obter mais detalhes sobre como a disponibilidade e a resiliência são usadas para medir a confiabilidade, consulte o Pilar de Confiabilidade do Well-Architected AWS Framework.

Perguntas chave

Disponibilidade

A disponibilidade é a porcentagem de tempo durante a qual uma workload está disponível para uso. As metas comuns incluem 99% (3,65 dias de inatividade permitidos por ano), 99,9% (8,77 horas) e 99,99% (52,6 minutos), com uma abreviação do número de nove na porcentagem (“dois noves” para 99%, “três noves” para 99,9% e assim por diante). A disponibilidade da solução de rede entre AWS e o data center local pode ser diferente da disponibilidade geral da solução ou do aplicativo.

As principais perguntas sobre a disponibilidade de uma solução de rede incluem:

- Meus AWS recursos podem continuar operando se não conseguirem se comunicar com meus recursos locais? Vice-versa?
- Devo considerar o tempo de inatividade programado para manutenção planejada como incluído ou excluído da métrica de disponibilidade?
- Como vou medir a disponibilidade da camada de rede, separada da integridade geral do aplicativo?

A seção Disponibilidade do Well-Architected Framework Reliability Pillar tem sugestões e fórmulas para a disponibilidade do cálculo.

Resiliência

Resiliência é a capacidade de uma worload se recuperar de interrupções na infraestrutura ou nos serviços, adquirir dinamicamente recursos de computação para atender à demanda e mitigar interrupções, como configurações incorretas ou problemas transitórios de rede. Se um componente

de rede redundante (link, dispositivos de rede, etc.) não tiver disponibilidade suficiente para fornecer sozinho a função esperada, ele terá baixa resiliência a falhas. A consequência é uma experiência de usuário ruim e degradada.

As principais perguntas para a resiliência de uma solução de rede incluem:

- Quantas falhas simultâneas e discretas devo permitir?
- Como posso reduzir pontos únicos de falha com as soluções de conectividade e com minha rede interna?
- Qual é a minha vulnerabilidade a eventos de negação distribuída de serviço (DDoS)?

Solução técnica

Primeiro, é importante observar que nem toda solução de conectividade de rede híbrida exige um alto nível de confiabilidade e que níveis crescentes de confiabilidade têm um aumento correspondente no custo. Em alguns cenários, um local primário pode exigir conexões confiáveis (redundantes e resilientes), pois o tempo de inatividade tem um impacto maior nos negócios, enquanto sites regionais podem não exigir o mesmo nível de confiabilidade devido ao menor impacto nos negócios em caso de falha. É recomendável consultar as Recomendações de AWS Direct Connect Resiliência, pois elas explicam as AWS melhores práticas para garantir alta resiliência com AWS Direct Connect o design.

Para obter uma solução confiável de conectividade de rede híbrida no contexto da resiliência, o design precisa levar em consideração os seguintes aspectos:

- Redundância: tente eliminar qualquer ponto único de falha no caminho de conectividade da rede híbrida, incluindo, mas não se limitando a, conexões de rede, dispositivos de rede de borda, redundância entre zonas de disponibilidade e locais de DX e fontes de alimentação de dispositivos, caminhos de fibra e sistemas operacionais. Regiões da AWS Para a finalidade e o escopo deste whitepaper, a redundância se concentra nas conexões de rede, nos dispositivos de borda (por exemplo, dispositivos de gateway do cliente), na localização do AWS DX e Regiões da AWS (para arquiteturas multirregionais).
- Componentes de failover confiáveis: em alguns cenários, um sistema pode estar funcionando, mas não estar executando suas funções no nível exigido. Essa situação é comum durante um único evento de falha em que se descobre que os componentes redundantes planejados estavam operando de forma não redundante. Sua carga de rede não tem outro lugar devido ao uso, o que resulta em capacidade insuficiente para toda a solução.

- Tempo de failover: o tempo de failover é o tempo necessário para que um componente secundário assuma totalmente a função do componente primário. O tempo de failover tem vários fatores: quanto tempo é necessário para detectar a falha, quanto tempo para ativar a conectividade secundária e quanto tempo para notificar o restante da rede sobre a alteração. A detecção de falhas pode ser aprimorada usando o Dead Peer Detection (DPD) para links VPN e o Bidirectional Forwarding Detection (BFD) para links. AWS Direct Connect O tempo para ativar a conectividade secundária pode ser muito baixo (se essas conexões estiverem sempre ativas), pode ser uma janela de tempo curta (se uma conexão VPN pré-configurada precisar ser ativada) ou maior (se os recursos físicos precisarem ser movidos ou novos recursos configurados). A notificação do restante da rede geralmente ocorre por meio de protocolos de roteamento dentro da rede do cliente, cada um com diferentes tempos de convergência e opções de configuração. A configuração desses protocolos está fora do escopo deste whitepaper.
- Engenharia de tráfego: a engenharia de tráfego no contexto do projeto resiliente de conectividade de rede híbrida visa abordar como o tráfego deve fluir por várias conexões disponíveis em cenários normais e de falha. É recomendável seguir o conceito de design para falhas, no qual é necessário verificar como a solução funcionará em diferentes cenários de falha e se ela será aceitável pela empresa ou não. Esta seção discute alguns dos casos de uso comuns de engenharia de tráfego que visam aprimorar o nível geral de resiliência da solução de conectividade de rede híbrida. A seção do AWS Direct Connect sobre roteamento e BGP fala sobre várias opções de engenharia de tráfego para influenciar o fluxo de tráfego (comunidades, preferência local do BGP, comprimento do caminho AS). Para projetar uma solução eficaz de engenharia de tráfego, você precisa ter uma boa compreensão de como cada um dos componentes de AWS rede lida com o roteamento IP em termos de avaliação e seleção de rotas, bem como os possíveis mecanismos para influenciar a seleção da rota. Os detalhes sobre isso estão fora do escopo deste documento. Para obter mais informações, consulte Ordem de avaliação da rota do gateway de trânsito, Prioridade de rota VPN Site-to-Site e a documentação do Direct Connect Routing e do BGP, conforme necessário.

> **Note**
>
> Na tabela de rotas da VPC, você pode fazer referência a uma lista de prefixos que tem regras adicionais de seleção de rotas. Para obter mais informações sobre esse caso de uso, consulte prioridade de rota para listas de prefixos. AWS Transit Gateway as tabelas de rotas também oferecem suporte a listas de prefixos, mas, uma vez aplicadas, elas são expandidas para entradas de rotas específicas.

Exemplo de conexões do Site-to-Site VPN com rotas mais específicas

Esse cenário é baseado em um pequeno site on-premises conectado a uma única Região da AWS por meio de conexões VPN redundante via Internet ao AWS Transit Gateway. O projeto de engenharia de tráfego descrito na Figura 10 mostra que, com a engenharia de tráfego, você pode influenciar a seleção do caminho que aumenta a confiabilidade da solução de conectividade híbrida da da seguinte forma:

- Conectividade híbrida resiliente: as conexões VPN redundantes fornecem a mesma capacidade de desempenho, suportam failover automatizado usando o protocolo de roteamento dinâmico (BGP) e aceleram a detecção de falhas de conexão usando a detecção de ponto morto de VPN.
- Eficiência de desempenho: configurar o ECMP em ambas as conexões VPN ao AWS Transit Gateway ajuda a maximizar a largura de banda geral da conexão VPN. Como alternativa, ao anunciar rotas diferentes e mais específicas junto com a rota resumida do site, a carga pode ser gerenciada de forma independente nas duas conexões VPN

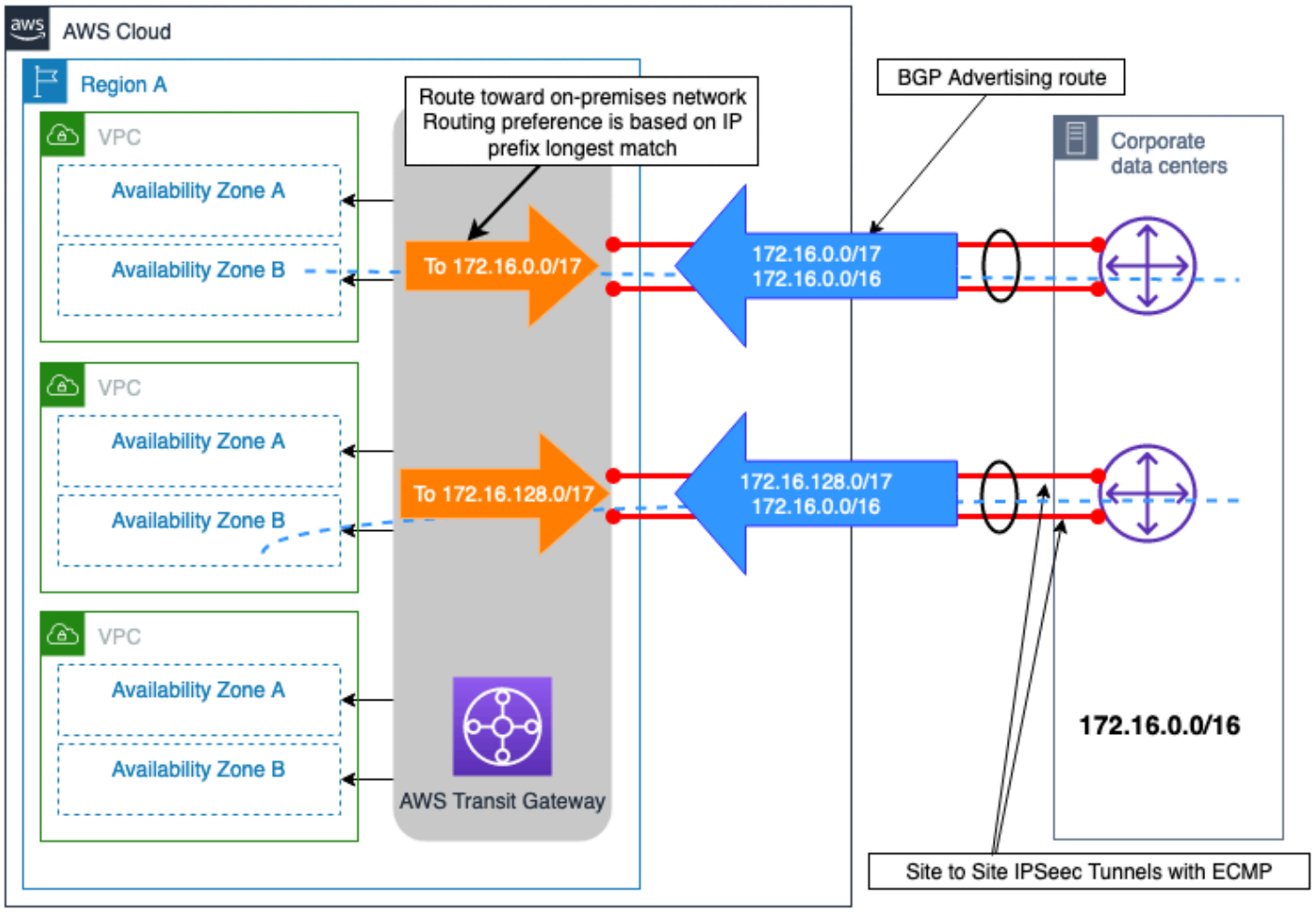

Figura 10 – conexões VPN site a site duplas com exemplo de rotas mais específicas

Exemplo de sites on-premises duplos com várias conexões DX

O cenário ilustrado na Figura 11 mostra dois sites de data center locais localizados em diferentes regiões geográficas e conectados AWS usando o modelo de conectividade de resiliência máxima (descrito nas Recomendações de resiliência) usando AWS Direct Connect com AWS Direct Connect DXGW e VGW. Esses dois sites on-premises estão interconectados entre si por meio de um link de interconexão de datacenter (DCI). Os prefixos IP locais (192.168.0.0/16) que pertencem às filiais remotas são anunciados nos dois sites do datacenter local. O caminho principal para esse prefixo deve ser o datacenter 1. O tráfego de e para as filiais remotas será transferido para o datacenter 2 em caso de falha do datacenter 1 ou de ambos os locais de DX. Além disso, há um prefixo IP específico do site para cada datacenter. Esses prefixos precisam ser acessados diretamente e por meio do outro local do datacenter em caso de falha em ambos os locais do DX.

Ao associar os atributos da comunidade BGP às rotas anunciadas para o AWS DXGW, você pode influenciar a seleção do caminho de saída do lado do DXGW. AWS Esses atributos da comunidade controlam AWS o atributo de preferência local do BGP atribuído à rota anunciada. Para obter mais informações, consulte Políticas de roteamento AWS DX e comunidades BGP.

Para maximizar a confiabilidade da conectividade no Região da AWS nível, cada par de conexões AWS DX configura o ECMP para que ambas possam ser utilizadas ao mesmo tempo na transferência de dados entre cada site local e. AWS

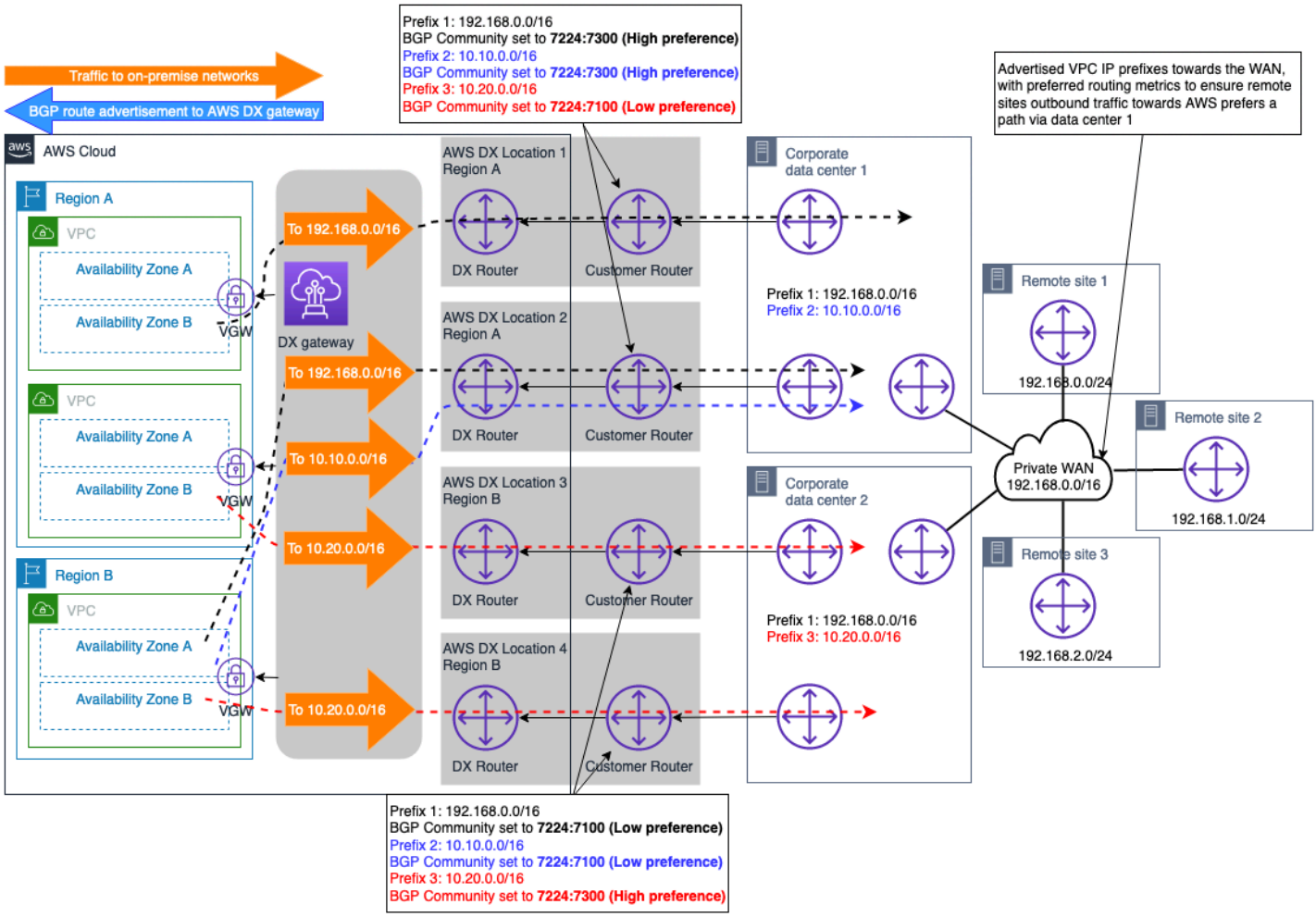

Figura 11 – exemplo de sites on-premises duplos com várias conexões DX

Com esse design, os fluxos de tráfego destinados às redes on-premises (com o mesmo tamanho de prefixo anunciado e comunidade BGP) serão distribuídos pelas conexões DX duplas por site usando ECMP. No entanto, se o ECMP não for necessário em toda a conexão DX, o mesmo conceito

discutido anteriormente e descrito na documentação Políticas de roteamento e Comunidades BGP poderá ser usado para projetar ainda mais a seleção do caminho em um nível de conexão DX.

Nota: se houver dispositivos de segurança no caminho dentro dos datacenters locais, esses dispositivos precisam ser configurados para permitir fluxos de tráfego saindo de um link DX e vindo de outro link DX (ambos os links utilizados com ECMP) no mesmo local do datacenter.

Exemplo de conexão VPN como backup para conexão AWS DX

A VPN pode ser selecionada para fornecer uma conexão de rede de backup a uma conexão AWS Direct Connect . Geralmente, esse tipo de modelo de conectividade é impulsionado pelo custo, pois oferece um nível mais baixo de confiabilidade para a solução de conectividade híbrida devido ao desempenho indeterminístico sobre a Internet, e não há SLA que possa ser obtido para uma conexão sobre a Internet pública. É um modelo de conectividade válido e econômico e deve ser usado quando o custo é a principal consideração prioritária e há um orçamento limitado, ou possivelmente como uma solução provisória até que um DX secundário possa ser provisionado. A Figura 12 ilustra o design desse modelo de conectividade. Uma consideração importante com esse design, em que as conexões VPN e DX terminam no AWS Transit Gateway, é que a conexão VPN pode anunciar um número maior de rotas em comparação com aquelas que podem ser anunciadas por meio de uma conexão DX conectada. AWS Transit Gateway Isso pode causar uma situação de roteamento abaixo do ideal. Uma opção para resolver esse problema é configurar a filtragem de rotas no dispositivo de gateway do cliente (CGW) para as rotas recebidas da conexão VPN, permitindo que somente as rotas resumidas sejam aceitas.

Nota: Para criar a rota resumida no AWS Transit Gateway, você precisa especificar uma rota estática para um anexo arbitrário na tabela de AWS Transit Gateway rotas para que o resumo seja enviado ao longo da rota mais específica.

Do ponto de vista da tabela de AWS Transit Gateway roteamento, as rotas para o prefixo local são recebidas da conexão AWS DX (via DXGW) e da VPN, com o mesmo tamanho de prefixo. Seguindo a lógica de prioridade de rota de AWS Transit Gateway, as rotas recebidas pelo Direct Connect têm uma preferência maior do que as recebidas pela VPN Site-to-Site e, portanto, o caminho até elas será o preferido para alcançar a (s) AWS Direct Connect rede (s) local (s).

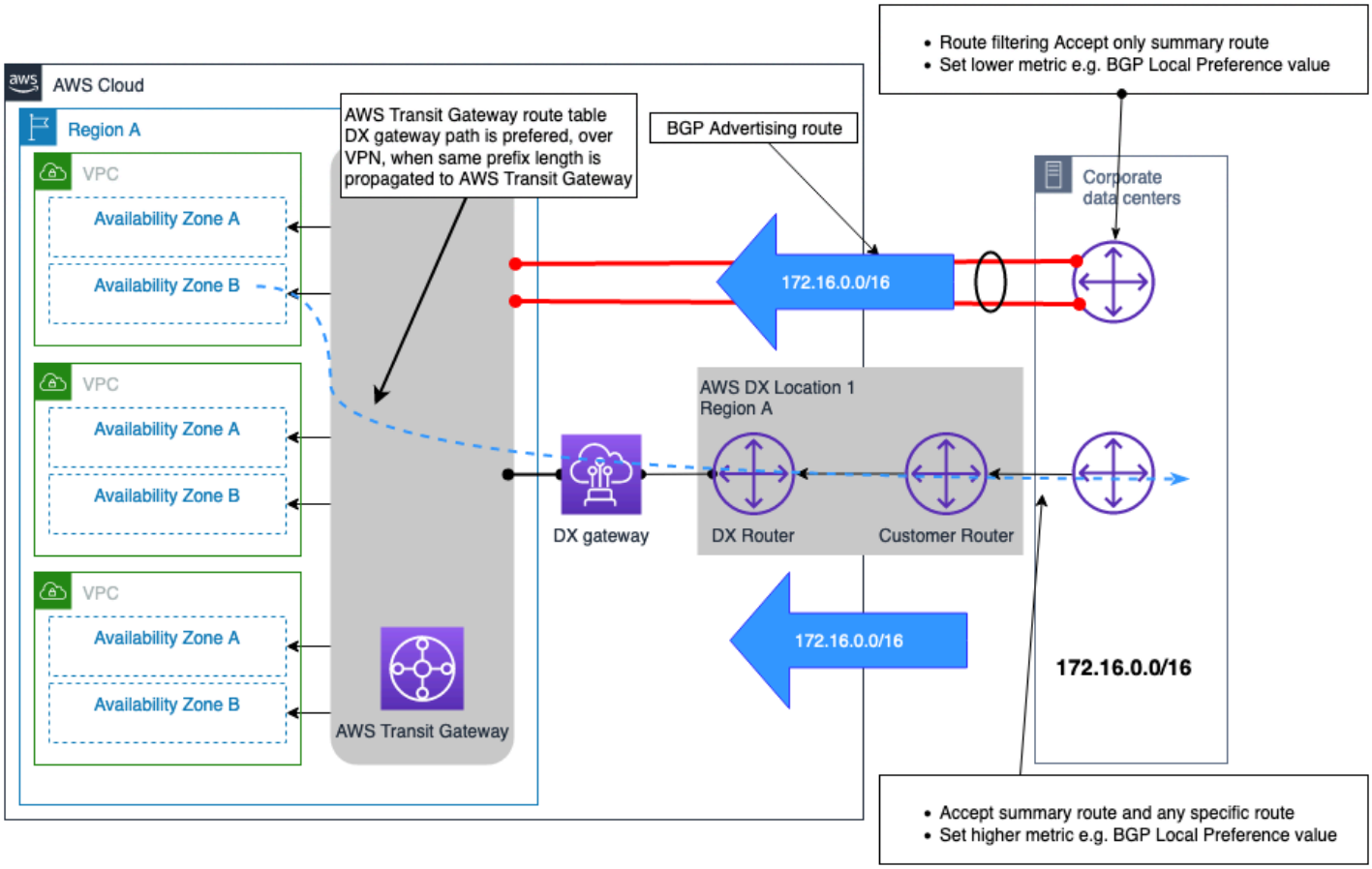

Figura 12 — Exemplo de conexão VPN como backup para conexão AWS DX

A árvore decisória a seguir orienta você na tomada da decisão desejada para obter uma conectividade de rede híbrida resiliente (o que resultará em uma conectividade de rede híbrida confiável). Para obter mais informações, consulte o AWS Direct Connect Resiliency Toolkit.

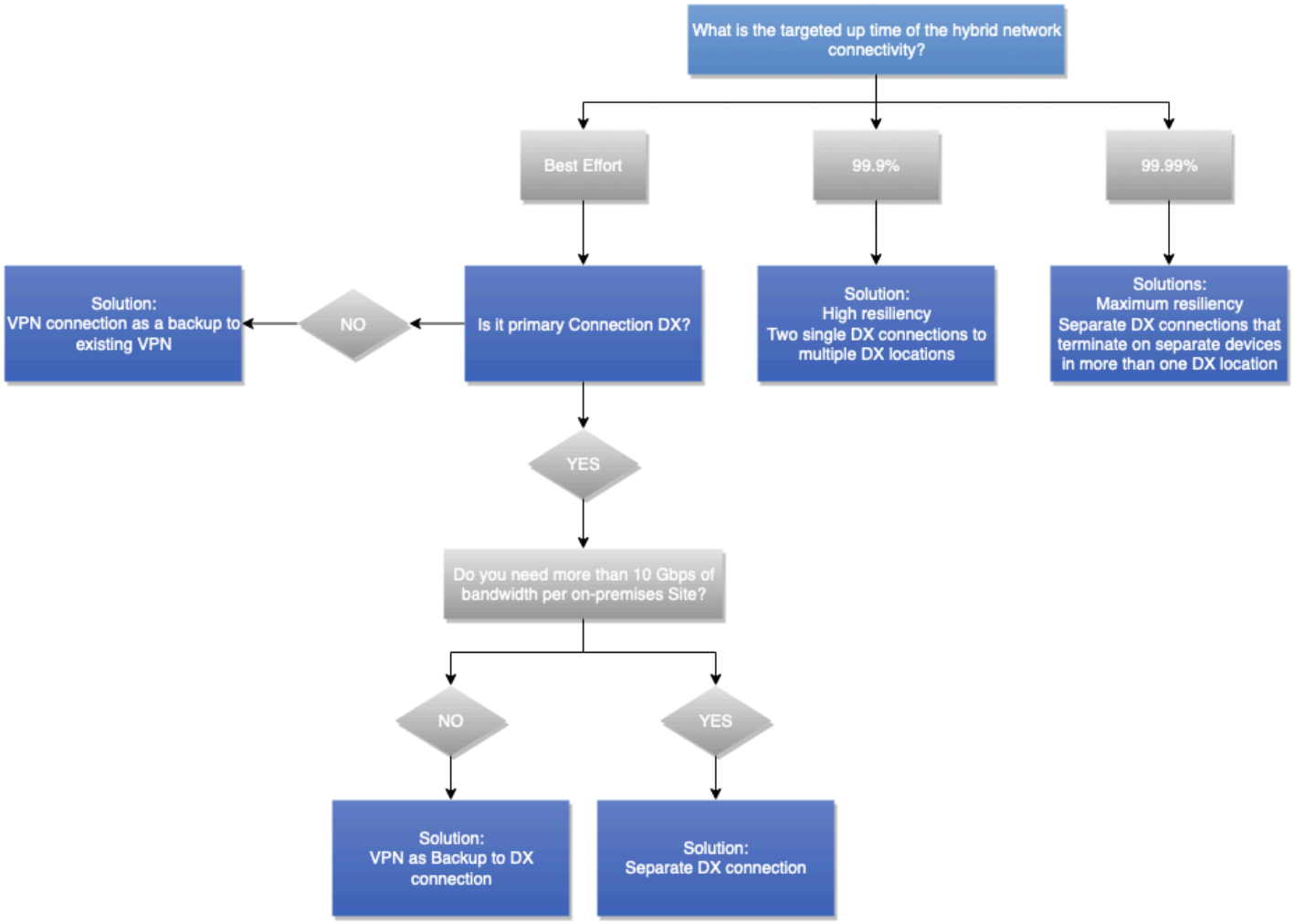

Figura 13 – Árvore de decisão de confiabilidade

VPN e SD-WAN gerenciadas pelo cliente

Definição

A conectividade com a Internet é uma mercadoria e a largura de banda disponível continua aumentando a cada ano. Alguns clientes optam por criar uma WAN virtual na Internet em vez de criar e operar uma WAN privada. Uma rede de área ampla definida por software (SD-WAN) permite que as empresas provisionem e gerenciem rapidamente de forma centralizada essa WAN virtual por meio do uso inteligente de software. Outros clientes optam por adotar VPNs autogerenciadas tradicionais de site para site.

Impacto nas decisões de design

A SD-WAN e as VPNs gerenciadas pelo cliente podem ser executadas pela Internet ou pelo AWS Direct Connect. A SD-WAN (ou qualquer sobreposição de VPN de software) é tão confiável quanto o transporte de rede subjacente. Portanto, as considerações sobre confiabilidade e SLA discutidas anteriormente neste whitepaper são aplicáveis aqui. Por exemplo, criar uma sobreposição de SD-WAN pela Internet não oferecerá a mesma confiabilidade do que se fosse construída sobre um AWS Direct Connect.

Definição de requisito

- Você usa o SD-WAN na rede on-premises?
- Você precisa de recursos específicos que só estão disponíveis em determinados dispositivos virtuais usados para terminação de VPN?

Soluções técnicas

AWS recomenda integrar a SD-WAN com AWS Transit Gateway a SD-WAN e publica uma lista dos fornecedores que oferecem suporte à integração. AWS Transit Gateway AWS pode atuar como um hub para sites SD-WAN ou como um site de fala. O AWS backbone pode ser usado para conectar diferentes hubs SD-WAN implantados em uma rede altamente confiável e AWS de alto desempenho. As soluções SD-WAN oferecem suporte a failover automatizado por meio de qualquer caminho disponível, monitoramento adicional e recursos de observabilidade em um único painel de gerenciamento. O uso extensivo de configuração e automação automáticas permite provisionamento e visibilidade rápidos em comparação com as WANs tradicionais. No entanto, o uso de encapsulamento e sobrecargas de criptografia não se compara aos links de fibra dedicados de alta velocidade usados na conectividade privada.

Em alguns casos, você pode optar por usar um dispositivo virtual com recurso de VPN. Os motivos para escolher um dispositivo virtual autogerenciado incluem recursos técnicos e compatibilidade com o resto da sua rede. Quando você seleciona uma VPN autogerenciada ou uma solução SD-WAN que usa um dispositivo virtual implantado em uma instância do EC2, você é responsável pelo gerenciamento desse dispositivo. Você também é responsável pela alta disponibilidade e pelo failover entre dispositivos virtuais. Esse design aumenta sua responsabilidade operacional; no entanto, pode fornecer mais flexibilidade. Os recursos e capacidades da solução dependem do dispositivo virtual selecionado.

AWS Marketplace contém muitos dispositivos virtuais de VPN que os clientes podem implantar no Amazon EC2. AWS recomenda começar com a VPN S2S AWS gerenciada e procurar outras opções se ela não atender aos seus requisitos. A sobrecarga de gerenciamento dos dispositivos virtuais é de responsabilidade do cliente.

Exemplo de caso de uso da Corp. Automotive

Esta seção do whitepaper demonstra como as considerações, as questões de definição de requisitos e as árvores de decisão são usadas para ajudá-lo a decidir sobre o design ideal da rede híbrida. Identificar e capturar os requisitos é importante, pois eles são usados como entrada para as árvores de decisão. A captura antecipada dos requisitos evita novas iterações de design. A interrupção total de um projeto se o design precisar ser revisitado e a retenção de recursos valiosos pode ser minimizada e, idealmente, evitada quando os requisitos são entendidos de antemão.

O exemplo da Corp. Automotive será usada em toda esta seção como cliente ilustrativo. Eles pretendem implantar inicialmente seu primeiro projeto de análise na AWS. O projeto de análise está focado na análise de dados de carros fabricados pela empresa e outros conjuntos de dados que já existem nos datacenters da empresa. Inicialmente, o grupo de arquitetura da empresa acha que precisará de uma Conta da AWS, uma Amazon VPC e algumas sub-redes para hospedar ambientes de produção e desenvolvimento. A equipe do projeto está ansiosa para começar e solicitou acesso ao ambiente de desenvolvimento o mais rápido possível. Eles pretendem entrar em produção daqui a três meses.

Exemplo: A Corp. Automotive também planeja usar a AWS em vários projetos adicionais, como a migração de seus sistemas ERP, Virtual Desktop Infrastructure (VDI) e outros 20 aplicativos on-premises para a AWS nos próximos 6 meses. Alguns requisitos para projetos adicionais ainda estão sendo definidos, mas está claro que seu uso da Nuvem AWS vai crescer.

A equipe de arquitetura decidiu aproveitar a abordagem descrita neste whitepaper. Eles usaram as questões de definição de requisitos descritas em cada consideração para capturar as informações para tomar suas decisões de design.

Eles começam com requisitos relacionados ao tipo de conectividade resumidos na tabela a seguir.

Tabela 4 – Exemplo de entradas de confiabilidade da corporação automotiva

Considerações sobre seleção do tipo de conectividade	Perguntas sobre a definição do requisito	Respostas
Tempo para implantação	Qual é o cronograma necessário para a implantação? Horas, dias, semanas ou meses?	• Desenvolvimento/teste: 1 mês • Produção: 3 meses

Considerações sobre seleção do tipo de conectividade	Perguntas sobre a definição do requisito	Respostas
Segurança	Seus requisitos e políticas de segurança permitem o uso de conexões criptografadas pela Internet para conexão com a AWS, ou exigem o uso de conexões de rede privadas?	• Desenvolvimento/teste: Site-to-Site VPN aceitável • Produção: Rede privada necessária
	Ao aproveitar as conexões de rede privada, a camada de rede precisa fornecer criptografia em trânsito?	Não, a criptografia da camada de aplicativo será usada.
SLA	É necessário um SLA de conectividade híbrida com créditos de serviço?	• Desenvolvimento/teste: Não • Produção: Sim
	Qual é a meta de tempo de atividade?	• Desenvolvimento/teste: N/D • Produção: 99,99%
	Toda a rede híbrida precisa cumprir uma meta de tempo de atividade?	• Desenvolvimento/teste: N/D • Produção: Sim
Desempenho	Qual é a taxa de transferência necessária?	• Desenvolvimento/teste: 100 Mbps • Produção: 500 Mbps crescendo para 2 Gbps
	Qual é a latência máxima aceitável entre uma rede da AWS e uma rede on-premises?	• Desenvolvimento/teste: Sem requisitos rígidos • Produção: Menos de 30 ms

Considerações sobre seleção do tipo de conectividade	Perguntas sobre a definição do requisito	Respostas
	Qual é a instabilidade de rede máxima aceitável?	• Desenvolvimento/teste: Sem requisitos rígidos • Produção: Jitter mínima necessária
Custos	Quantos dados você enviaria para a AWS por mês?	• Desenvolvimento/teste: 2 TB • Produção: 20 TB crescendo para 50 TB
	Quantos dados você enviaria da AWS por mês?	• Desenvolvimento/teste: 1 TB • Produção: 10 TB crescendo para 25 TB
	Essa conectividade é permanente?	Sim

Com base nos requisitos recebidos, a equipe de arquitetura seguiu a árvore decisória do tipo de conectividade da Figura 9. Isso permitiu que a equipe de arquitetura decidisse sobre o tipo de conectividade para os ambientes de desenvolvimento, teste e produção. Para o ambiente de produção, eles consideraram os requisitos imediatos e futuros. Para desenvolvimento e teste, a Corp. Automotive estabelecerá uma VPN site a site pela Internet. Para a produção, eles trabalharão com um provedor de serviços para conectar sua rede corporativa com o AWS Direct Connect. A Corp. Automotive inicialmente considerou usar uma Conexão hospedada do Direct Connect; no entanto, devido aos requisitos de um SLA específico da AWS, eles selecionaram conexões dedicadas do Direct Connect.

Depois de decidir sobre o tipo de conectividade, a próxima etapa é capturar os requisitos que afetam a seleção do design de conectividade. Isso está relacionado ao design lógico, como as conexões são configuradas e quais serviços da AWS usar para dar suporte aos requisitos comerciais e técnicos.

Para capturar os requisitos do modelo de escalabilidade e comunicação, a equipe de arquitetura usou as perguntas de definição de requisitos das seções associadas deste whitepaper. Os requisitos relacionados a essas duas considerações estão resumidos na tabela a seguir.

Tabela 5 – perguntas sobre a definição de requisitos

Considerações sobre seleção do design de conectividade	Perguntas sobre a definição do requisito	Respostas
Escalabilidade	Qual é o número atual ou previsto de VPCs que exigem conectividade com sites on-premises?	2 inicialmente, crescendo para 30 em 6 meses
	As VPCs são implantadas em uma Região da AWS única ou em várias regiões?	Região única
	Quantos sites on-premises precisam estar conectados à AWS?	2 datacenters
	Quantos dispositivos de gateway do cliente você tem por site que precisam ser conectados à AWS?	2 roteadores por datacenter
	Quantas rotas espera-se serem anunciadas para as AWS VPCs e qual é o número de rotas esperadas a serem recebidas do lado da AWS?	• Rotas a serem anunciadas para a AWS: 20 rotas • Rotas a serem recebidas da AWS: 1 rota de /16
	Existe algum plano para considerar o aumento da largura de banda da conexão com a AWS em um futuro próximo?	• Desenvolvimento/teste: 100 Mbps • Produção: 500 Mbps crescendo para 2 Gbps.

Considerações sobre seleção do design de conectividade	Perguntas sobre a definição do requisito	Respostas
Modelos de design de conectividade	Há um requisito para que a comunicação entre VPCs seja habilitada (dentro de uma região e/ou entre regiões)?	Sim, dentro de uma Região da AWS
	Há algum requisito para acessar serviços de endpoints públicos da AWS diretamente on-premises?	Sim
	Há um requisito para acessar serviços da AWS usando endpoints da VPC on-premises?	Não

Com base nas entradas, a equipe de arquitetura seguiu a árvore de decisão da seção Design de conectividade. Depois de prever que o número de VPCs crescerá de 2 para 30 nos próximos 6 meses, a equipe de arquitetura decidiu usar o AWS Transit Gateway como gateway de terminação para a conexão e para o roteamento entre VPCs. AWS Transit Gateways independentes encerrarão a conexão VPN usada para desenvolvimento e teste, e para a conectividade de produção com o AWS Direct Connect. O uso de AWS Transit Gateways separados simplifica o gerenciamento de mudanças e fornece uma demarcação clara entre os ambientes de desenvolvimento/teste e produção. Para a produção, o gateway de AWS Direct Connect é necessário por causa do AWS Transit Gateway. Um VIF público será usado para acessar serviços de endpoint públicos da AWS. A Figura 14 ilustra o caminho percorrido na árvore de decisão com base nos requisitos coletados.

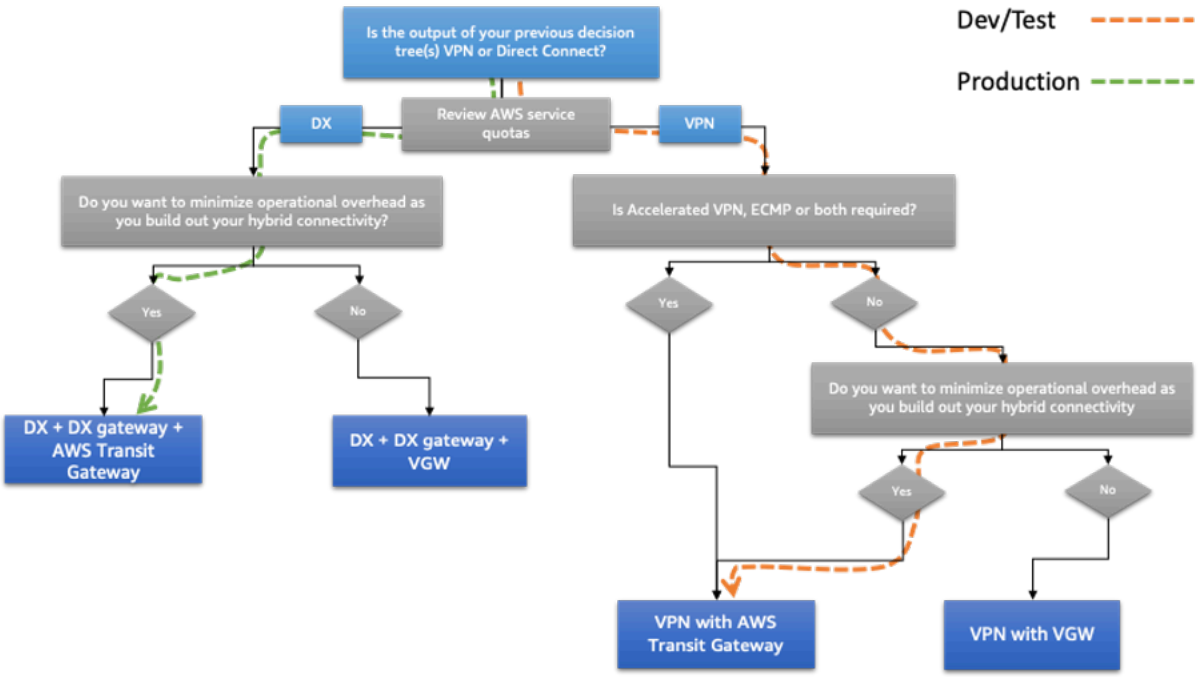

Figura 14 – Árvore de decisão do design de conexão da Corp. Automotive

Depois de decidir sobre a solução para atender aos requisitos do modelo de escalabilidade e comunicação, a próxima etapa é capturar os requisitos associados à confiabilidade. Isso está relacionado ao nível exigido de disponibilidade e resiliência.

Para capturar os requisitos confiabilidade, a equipe de arquitetura usou as perguntas de definição de requisitos da seção associada deste whitepaper. Os requisitos estão resumidos na tabela a seguir.

Tabela 6 – Perguntas sobre requisitos de confiabilidade

Considerações sobre seleção do design de conectividade	Perguntas sobre a definição do requisito	Respostas
Confiabilidade	Qual é a magnitude do impacto nos negócios em caso de falha de conectividade com a AWS?	• Desenvolvimento/teste: Baixa • Produção: Alta
	Do ponto de vista comercial, o custo de acompanhar uma falha de conectividade com a AWS supera o custo da	• Desenvolvimento/teste: Não • Produção: Sim

Considerações sobre seleção do design de conectividade	Perguntas sobre a definição do requisito	Respostas
	implantação de um modelo de conectividade altamente confiável para a AWS?	

Com base nas informações recebidas, a equipe de arquitetura seguiu a árvore de decisão das seções de considerações de confiabilidade abordadas anteriormente neste whitepaper. Depois de considerar a meta de tempo de atividade de 99,99% para a conectividade de produção e o alto impacto nos negócios se houvesse uma interrupção do serviço, a equipe de arquitetura decidiu usar 2 locais do Direct Connect e ter 2 links de cada data center local para cada local do Direct Connect (4 links no total). A conectividade VPN usada para desenvolvimento e teste também usará duas conexões VPN para redundância adicional. Usando as técnicas de engenharia de rotas discutidas na seção de confiabilidade, a conectividade será configurada da seguinte forma:

- Para desenvolvimento e teste, a carga do tráfego será balanceada usando ECMP nos 2 túneis que vão para o datacenter primário. Isso permite uma maior throughput. Os túneis que vão para o data center secundário serão usados em caso de falha dos túneis primários.
- Para produção, a latência entre on-premises e AWS em qualquer um dos locais do Direct Connect é muito semelhante. Nesse caso, foi decidido balancear a carga do tráfego entre AWS e on-premises nas duas conexões que vão para o datacenter principal para os sistemas on-premises implantados no datacenter primário. Da mesma forma, para sistemas locais executados no datacenter secundário, a throughput será balanceada entre as duas conexões com o datacenter secundário. Em caso de falha nas conexões, o BGP facilitará um failover automático.

A Figura 15 ilustra o caminho percorrido na árvore de decisão com base nos requisitos coletados.

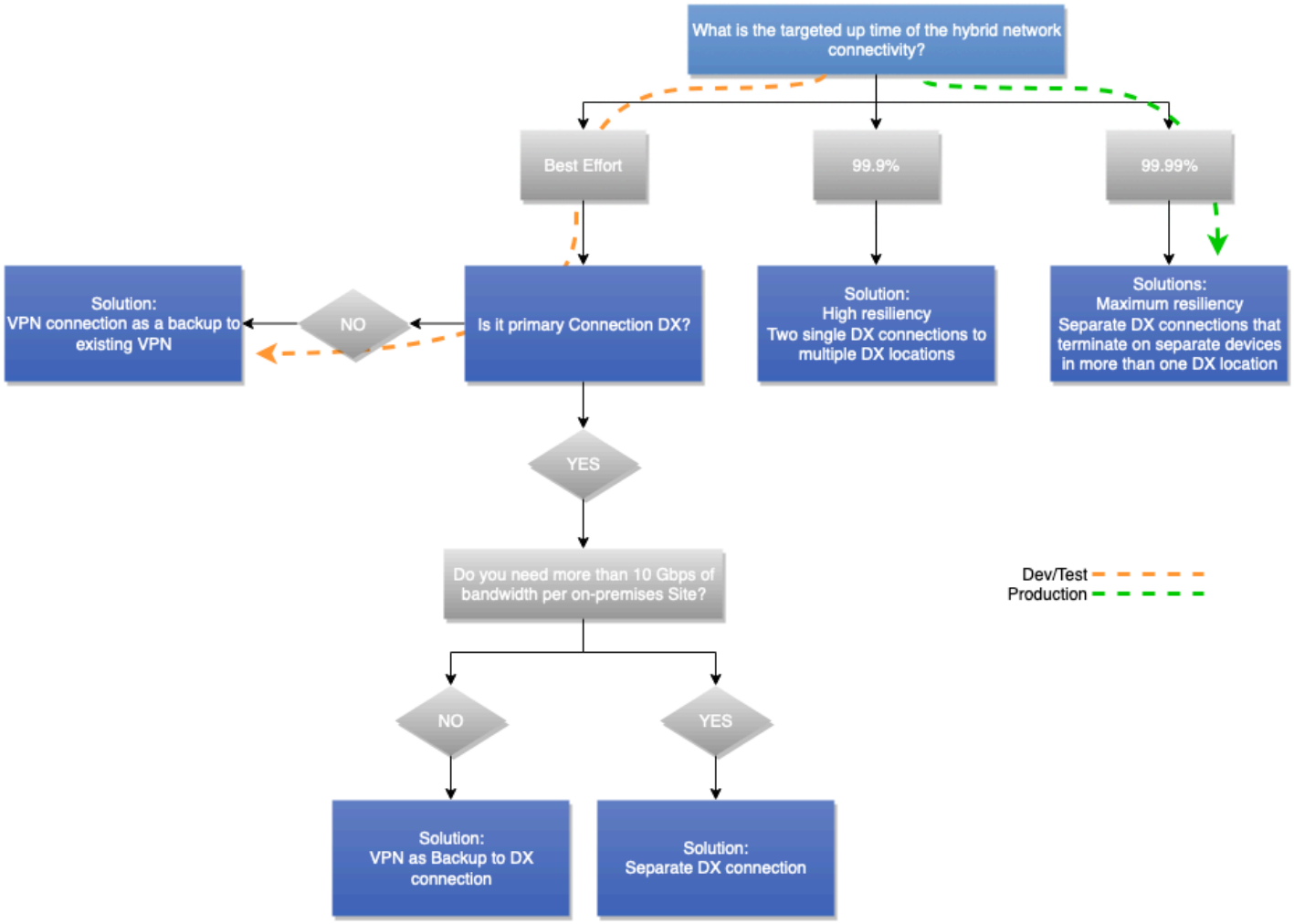

Figura 15 – Árvore de decisão de confiabilidade da Corp. Automotive

Arquitetura selecionada pela Corp. Automotive

O diagrama a seguir ilustra a arquitetura selecionada pela Corp. Automotive após coletar os requisitos e navegar pelas árvores decisórias abordadas nas seções anteriores deste whitepaper.

Ela usa o AWS S2S VPN pela Internet, terminando no AWS Transit Gateway para desenvolvimento e teste. Em seguida, ela usa o AWS Direct Connect com o gateway do Direct Connect e um segundo AWS Transit Gateway para o tráfego de produção. O AWS Transit Gateway é usado para roteamento entre VPCs. Do ponto de vista do caminho de dados, os túneis VPN para o datacenter primário são usados como caminhos primários para desenvolvimento e teste, com os túneis para o datacenter secundário usados como caminhos de failover. Para o tráfego de produção, todas as conexões são usadas simultaneamente. O tráfego da AWS prefere a conexão de rede mais opcional

com base no datacenter no qual o sistema local está localizado. A Corp. Automotive usa técnicas similares de engenharia de rotas para preferir o caminho apropriado quando o tráfego é enviado para a AWS, garantindo que caminhos de tráfego simétricos sejam usados para minimizar o uso da rede corporativa entre datacenters primários e secundários locais.

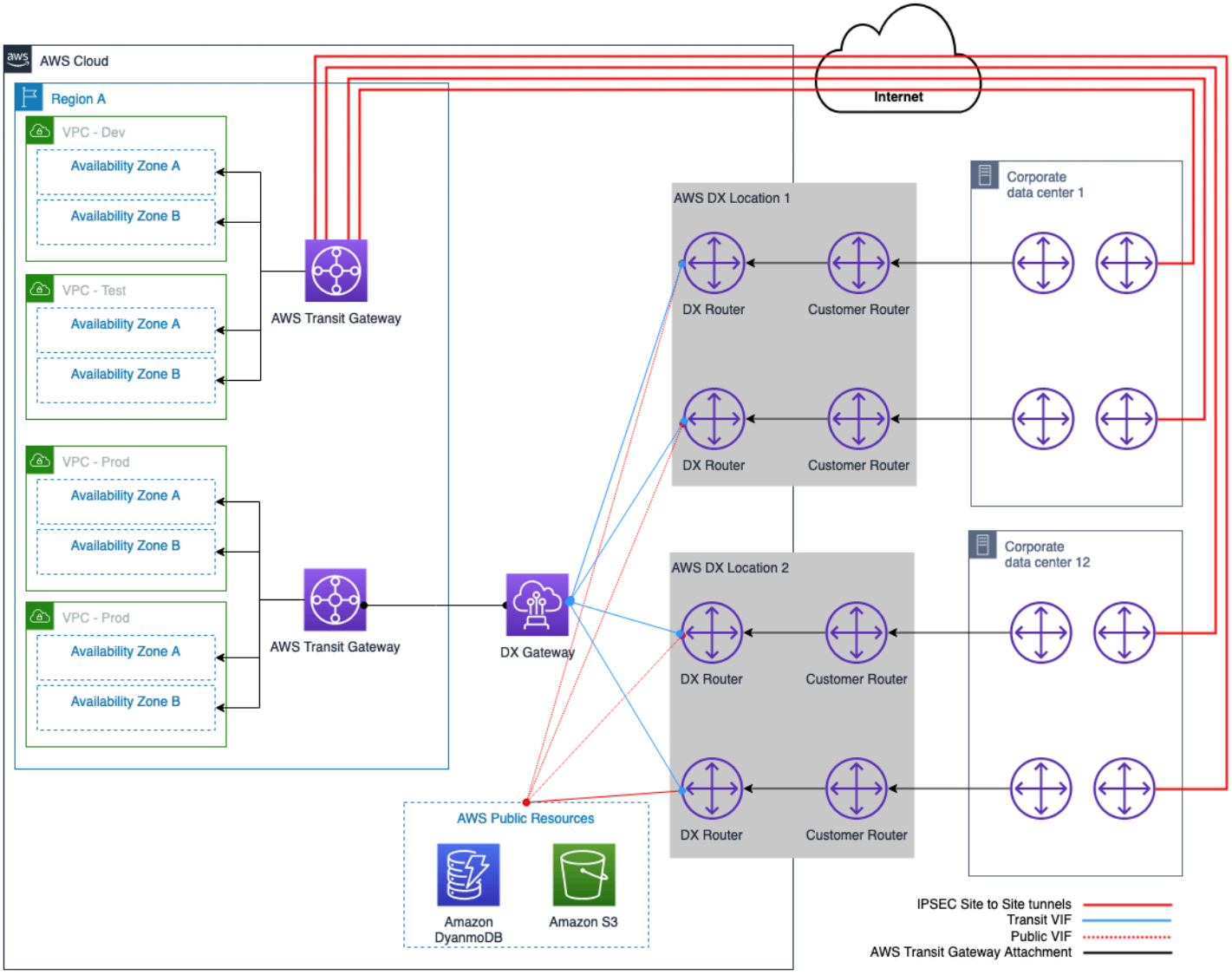

Figura 16 – Modelo de conectividade híbrida selecionado pela Corp. Automotive

Conclusão

Um modelo de conectividade híbrida é um dos pontos de partida fundamentais para a adoção da computação em nuvem. Uma rede híbrida pode ser construída com uma arquitetura ideal seguindo o processo de seleção do modelo de conectividade descrito neste whitepaper.

O processo consiste em considerações organizadas em uma ordem lógica. A ordem se assemelha muito a um modelo mental seguido por arquitetos experientes de rede e nuvem. Dentro de cada grupo de considerações, as árvores de decisão permitem a seleção rápida do modelo de conectividade, mesmo com requisitos de entrada limitados. É possível descobrir que algumas considerações e impactos correspondentes apontam para soluções diferentes. Nesses casos, como tomador de decisões, talvez você precise comprometer alguns requisitos e selecionar a solução ideal que atenda aos requisitos técnicos e comerciais.

Colaboradores

Os colaboradores deste documento incluem:

- James Devine, principal arquiteto de soluções, Amazon Web Services
- Andrew Gray, principal arquiteto de soluções, redes, Amazon Web Services
- Maks Khomutskyi, arquiteto sênior de soluções, Amazon Web Services
- Marwan Al Shawi, arquiteto de soluções, Amazon Web Services
- Santiago Freitas, diretor de tecnologia, Amazon Web Services
- Evgeny Vaganov, arquiteto especialista em soluções, redes, Amazon Web Services
- Tom Adamski, arquiteto especialista em soluções, redes, Amazon Web Services
- Armstrong Onaiwu, arquiteto de soluções, Amazon Web Services

Outras fontes de leitura

- Construção de uma infraestrutura de rede da AWS Multi-VPC escalável e segura
- Opções de DNS de nuvem híbrida para Amazon VPC
- Opções de conectividade da Amazon Virtual Private Cloud
- Documentação do Amazon Virtual Private Cloud
- Documentação do AWS Direct Connect
- Qual é a diferença entre uma interface virtual hospedada (VIF) e uma conexão hospedada?

Revisões do documento

Para ser notificado sobre atualizações desse whitepaper, inscreva-se no feed RSS.

Alteração	Descrição	Data
Atualização secundária	Atualizado para refletir o aumento do limite de cota de DX.	10 de julho de 2023
Atualização principal	Atualizado para incorporar as melhores práticas, serviços e recursos mais recentes.	6 de julho de 2023
Atualização secundária	Diagramas de arquitetura de referência atualizados para refletir as mudanças na cota de DX.	27 de junho de 2023
Atualização secundária	Links quebrados corrigidos.	22 de março de 2022
Publicação inicial	Whitepaper publicado pela primeira vez	22 de setembro de 2020

Avisos

Os clientes são responsáveis por fazer sua própria avaliação independente das informações contidas neste documento. Este documento: (a) é apenas para fins informativos, (b) representa as ofertas de produto e práticas da AWS, que estão sujeitas a alterações sem aviso prévio e (c) não cria nenhum compromisso ou garantia da AWS e de suas afiliadas, fornecedores ou licenciadores. Os produtos ou serviços da AWS são fornecidos “no estado em que se encontram”, sem garantias, representações ou condições de qualquer tipo, expressas ou implícitas. As responsabilidades e as obrigações da AWS com os seus clientes são controladas por contratos da AWS, e este documento não é parte, nem modifica, qualquer contrato entre a AWS e seus clientes.

AWS Glossário

Para obter a terminologia mais recente da AWS, consulte o glossário da AWS na Referência do Glossário da AWS.

As traduções são geradas por tradução automática. Em caso de conflito entre o conteúdo da tradução e da versão original em inglês, a versão em inglês prevalecerá.