SINGLE DATA CENTER HA NETWORK CONNECTIVITY

"How do I implement HA connectivity between my data center and my VPC?"

Overview

Amazon Web Services (AWS) offers customers the ability to achieve highly available network connections between Amazon Virtual Private Cloud (Amazon VPC) and their on-premises infrastructure. This capability extends customer access to AWS resources in a reliable, scalable, and cost-effective way. While many customers are able to quickly determine the type of connections they want to establish (either VPN or AWS Direct Connect[1]), they often struggle with understanding how to make these connections highly available and how to best leverage redundant connections. This document provides AWS customers with best practices, recommendations, and common configurations to help build highly available, efficient connections between AWS and a single remote data center. When possible, AWS recommends connecting from multiple data centers in order to add physical-location redundancy, in addition to hardware redundancy; however, we realize this might not be practical for all customers. For information about creating highly available network connections from multiple locations, please see the Multiple Data Center HA Network Connectivity Solution Brief.

The following sections are applicable to customers who have already determined the appropriate remote connectivity option for their use case,[1] and assume basic knowledge of Amazon VPC, Virtual Private Network connections, AWS Direct Connect, connecting and routing between remote networks, and dynamic routing protocols.

General Best Practices

Highly available, fault-tolerant network connections are key to a well-architected system. When designing remote connections, consider using redundant hardware and telecommunications providers on both sides of the network connection. Additionally, it is a best practice to use dynamically routed, active/active connections for automatic load balancing and failover across redundant network connections. Keeping these topology guidelines in mind, consider the following best practices when connecting to AWS:

- Leverage multiple dynamically routed, rather than statically routed, connections to AWS. This will allow remote connections to fail over automatically between redundant connections. Dynamic routing also enables remote connections to automatically leverage available preferred routes, if applicable, to the on-premises network.
- Avoid relying on a single on-premises device, even if you were planning to use multiple interfaces for availability. Highly available connections require redundant hardware, even when connecting from the same physical location.
- When selecting AWS Direct Connect network service providers, consider a dual-vendor approach, if financially feasible, to ensure private-network diversity.
- Provision sufficient network capacity to ensure that the failure of one network connection does not overwhelm and degrade redundant connections.

Application on the AWS Platform

The following sections provide high-availability options for redundant VPN connections, redundant AWS Direct Connect connections, and an AWS Direct Connect connection with a backup VPN connection. An organization's business-availability and application requirements will help determine the most appropriate configuration for each use case. For example, the most robust option is to implement multiple AWS Direct Connect connections with circuits from different telecommunication providers.

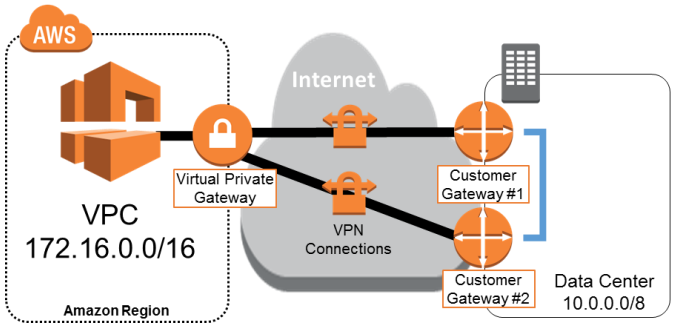
[1] Please see the *AWS Network Connectivity Options* whitepaper for more information.

Redundant Active/Active VPN Connections

Many AWS customers choose to implement VPN connections because they can be a quick, easy, and cost-effective way to set up remote connectivity to a VPC. To enable redundancy, each AWS Virtual Private Gateway (VGW) has two VPN endpoints[2] with capabilities for static and dynamic routing. Although statically routed VPN connections from a single customer gateway are sufficient for establishing remote connectivity to a VPC, this is not a highly available configuration. The best practice for making VPN connections highly available is to use redundant customer gateways and dynamic routing for automatic failover between AWS and customer VPN endpoints. For simplicity, the diagram in the next section depicts each VPN connection, consisting of two IPsec tunnels to both VGW endpoints, as a single line.

Configuration Details

The configuration in this example consists of four fully meshed, dynamically routed IPsec tunnels between both VGW endpoints and two customer gateways. Both AWS and the customer provide redundant hardware and use two independently configured VPN endpoints. In addition to leveraging the AWS-provided VPN configuration templates, customers must propagate VPN-connection availability into their internal network to ensure their systems will be able to utilize either path when necessary (depicted with the blue line in the diagram to the right). Most single data centers will not require additional control over network routing and path selection. See the Multiple Data Center HA Network Connectivity Solution Brief for route-manipulation options for this type of router configuration.



Considerations

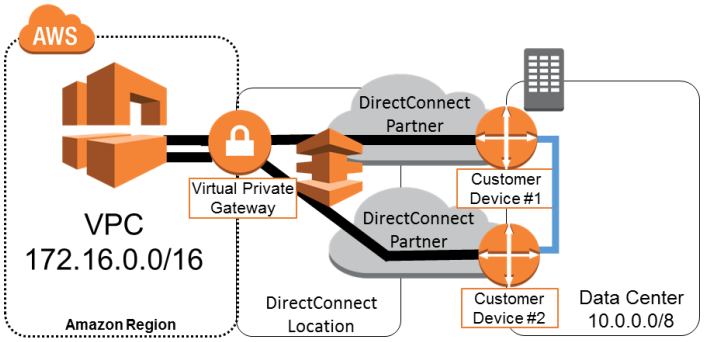
This configuration relies on the Internet to carry traffic between the on-premises network and VPC. Any location-specific disruption (e.g., loss of power or cable damage outside the facility) can still affect network connectivity to AWS. Additionally, although AWS leverages multiple Internet Service Providers (ISPs), and even if the customer leverages multiple ISPs, an Internet service disruption can still affect VPN network connectivity due to the interdependence of ISPs and Internet routing. The only way to control the exact network path of your traffic is to provision private connectivity with AWS Direct Connect (see the next option).

Redundant Active/Active AWS Direct Connect Connections

Many AWS customers establish private connectivity between AWS and their data center, office, or colocation environment with AWS Direct Connect to reduce network costs, increase bandwidth throughput, or provide a more consistent network experience than Internet-based connections. Because each dedicated, physical connection is in one AWS Direct Connect location, multiple dynamically routed AWS Direct Connect connections are necessary to achieve high availability. Architectures with the highest levels of availability will leverage different AWS Direct Connect partner networks to ensure network-provider redundancy. For simplicity, the diagram in the next section depicts each AWS Direct Connect connection, consisting of a physical connection that contains logical connections, as one continual line.[3]

Configuration Details

The configuration in this example consists of AWS Direct Connect connections to two separate AWS Direct Connect routers from two independently configured customer devices. In addition to leveraging the AWS-provided Direct Connect configuration templates, customers must propagate AWS Direct Connect connection availability into their internal network to ensure their systems will be able to utilize either path when necessary (depicted with the blue line in the diagram to the right). Most single data centers will not require additional control over network routing and path selection. However, the Multiple Data Center HA Network Connectivity Solution Brief has more details on route manipulation options for this type of router configuration.



[2] http://docs.aws.amazon.com/AmazonVPC/latest/UserGuide/VPC_VPN.html#VPNTunnels

[3] For more information on AWS Direct Connect physical and logical connections, please see http://docs.aws.amazon.com/directconnect/latest/UserGuide/Welcome.html

Considerations

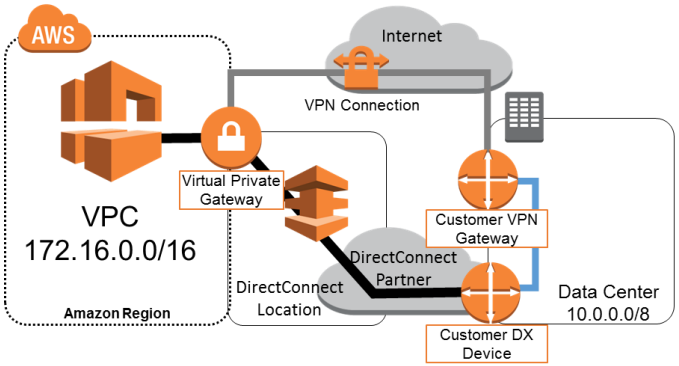
AWS Direct Connect allows you to create resilient connections because you have full control over the network path and network providers between your remote network and AWS. Choose network providers and AWS Direct Connect locations that align with your organization's risk tolerance, financial expectations, and data-center-connectivity policies. For example, if you are concerned about the risk associated with an individual network-provider outage, consider using a different network provider for each AWS Direct Connect connection. Likewise, leveraging different AWS Direct Connect locations (e.g., CoreSite and Equinix in US East) will reduce the risk that a facility failure will interrupt your network connectivity with AWS. However, because this design relies on a single customer location to provide connectivity to AWS, any location-specific disruption (e.g., loss of power or cable damage outside the facility) can still affect network connectivity to AWS, despite leveraging redundant network providers.

AWS Direct Connect with Backup VPN Connection

Some AWS customers would like the benefits of one or more AWS Direct Connect connections for their primary connectivity to AWS, coupled with a lower-cost backup connection. To achieve this objective, they can establish AWS Direct Connect connections with a VPN backup, as depicted in the diagram below.

Configuration Details

The configuration in this example consists of two dynamically routed connections, one using AWS Direct Connect and the other using a VPN connection from two different customer devices. AWS provides example router configurations to assist in establishing both AWS Direct Connect and dynamically routed VPN connections. By default, AWS will always prefer to send traffic over your AWS Direct Connect connection, so no additional AWS-specific configuration is required to define primary and backup connections. However, customers should configure AWS Direct Connect and VPN-specific internal-route propagation to ensure internal systems select the appropriate paths.[4]



Considerations

This approach allows you to choose the primary network path and network provider for your AWS traffic, with the option of using a different provider for a backup VPN connection. Choose network providers and AWS Direct Connect locations that align with your organization's risk tolerance, financial expectations, and data-center connectivity policies. For example, if you are concerned about the risk associated with an individual network-provider outage, consider different network providers for AWS Direct Connect and Internet connectivity. However, because this design relies on a single customer location to provide connectivity to AWS, any location-specific disruption (e.g., loss of power or cable cut outside the facility) can still affect network connectivity to AWS, despite leveraging redundant network providers. Additionally, make sure to monitor AWS Direct Connect utilization to ensure that a VPN connection will be a sufficient backup to support your application's latency and bandwidth requirements.

Resources

Amazon VPC Network Connectivity Options	https://aws.amazon.com/whitepapers/amazon-vpc-connectivity-options/ AWS whitepaper on common network connectivity options, including the integration of customer networks with Amazon VPC.
AWS Direct Connect Documentation	https://aws.amazon.com/documentation/direct-connect/
Amazon VPC Documentation	https://aws.amazon.com/documentation/vpc/
Multiple Data Center HA Network Connectivity	https://d0.awsstatic.com/aws-answers/AWS_Multiple_Data_Center_HA_Network_Connectivity.pdf AWS Solution Brief on options for creating highly available connections from multiple data centers to AWS.

[4] The Multiple Data Center HA Network Connectivity Solution Brief has more details on route manipulation options for this scenario; however, the default configuration is typically sufficient for most customers connecting to AWS from a single data center.