Interconnecting Networks

Priyanka Vergadia

Hi I'm Priyanka Vergadia, a developer advocate for Google Cloud. In this module, we focus on Interconnecting Networks.

Different applications and workloads require different network connectivity solutions. That is why Google supports multiple ways to connect your infrastructure to GCP.

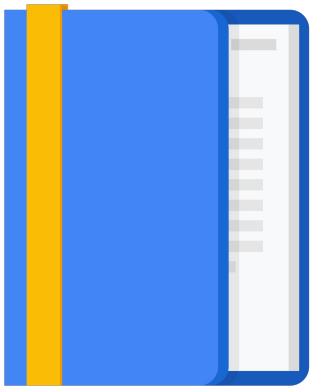
Agenda

Cloud VPN

Lab

Cloud Interconnect and Peering

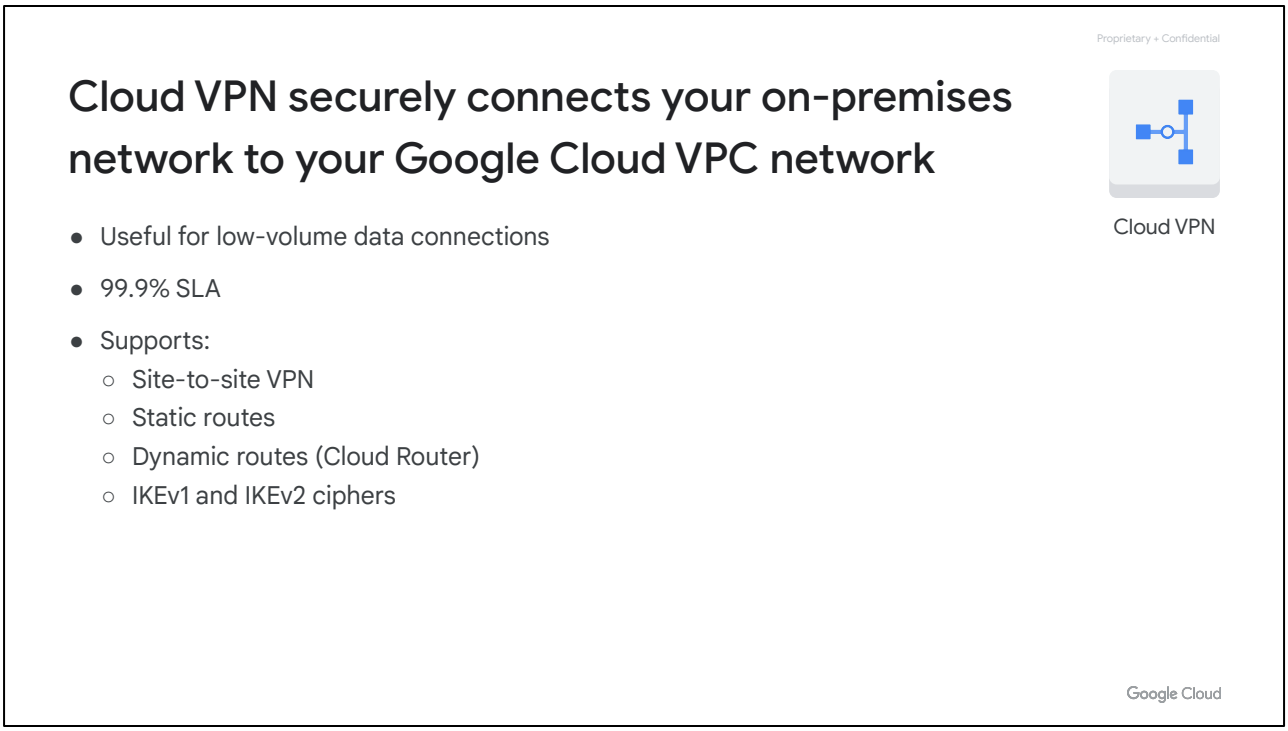
Sharing VPC Networks



Google Cloud

In this module, we'll focus on GCP's hybrid connectivity products, which are Cloud VPN, Cloud Interconnect, and Peering. We'll also look at options for sharing VPC networks within GCP.

Let's start by talking about Cloud VPN so that you can explore it in a lab.



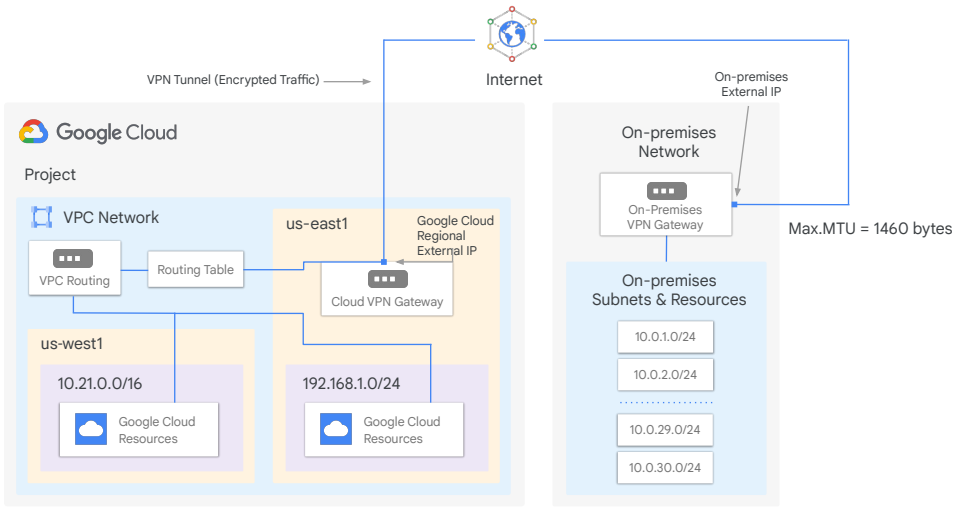
Cloud VPN securely connects your on-premises network to your Google Cloud VPC network through an IPsec VPN tunnel. Traffic traveling between the two networks is encrypted by one VPN gateway, then decrypted by the other VPN gateway. This protects your data as it travels over the public internet, and that's why Cloud VPN is useful for low-volume data connections.

As a managed service, Cloud VPN provides an SLA of 99.9% service availability and supports site-to-site VPN, static and dynamic routes, and IKEv1 and IKEv2 ciphers. Cloud VPN doesn't support use cases where client computers need to "dial in" to a VPN using client VPN software. Also, dynamic routes are configured with Cloud Router, which we will cover briefly.

For more information about the SLA and these features, please refer to the documentation link.

- Cloud VPN overview: https://cloud.google.com/vpn/docs/concepts/overview

Classic VPN topology



Let me walk through an example of Cloud VPN. This diagram shows a Classic VPN connection between your VPC and on-premises network. Your VPC network has subnets in us-east1 and us-west1, with Google Cloud resources in each of those regions. These resources are able to communicate using their internal IP addresses because routing within a network is automatically configured (assuming that firewall rules allow the communication).

Now, in order to connect to your on-premises network and its resources, you need to configure your Cloud VPN gateway, on-premises VPN gateway, and two VPN tunnels. The Cloud VPN gateway is a regional resource that uses a regional external IP address.

Your on-premises VPN gateway can be a physical device in your data center or a physical or software-based VPN offering in another cloud provider's network. This VPN gateway also has an external IP address.

A VPN tunnel then connects your VPN gateways and serves as the virtual medium through which encrypted traffic is passed. In order to create a connection between two VPN gateways, you must establish two VPN tunnels. Each tunnel defines the connection from the perspective of its gateway, and traffic can only pass when the pair of tunnels is established.

Now, one thing to remember when using Cloud VPN is that the maximum transmission unit, or MTU, for your on-premises VPN gateway cannot be greater than 1460 bytes. This is because of the encryption and encapsulation of packets. For more information about this MTU consideration, please refer to the documentation link in the course resources.

[**MTU considerations:**
https://cloud.google.com/vpn/docs/concepts/mtu-considerations]

HA VPN overview

- Provides 99.99% service availability.
- Google Cloud automatically chooses two external IP addresses.
  - Supports multiple tunnels
  - VPN tunnels connected to HA VPN gateways must use dynamic (BGP) routing
- Supports site-to-site VPN for different topologies/configuration scenarios:
  - An HA VPN gateway to peer VPN devices
  - An HA VPN gateway to an Amazon Web Services (AWS) virtual private gateway
  - Two HA VPN gateways connected to each other

In addition to Classic VPN, Google Cloud also offers a second type of Cloud VPN gateway, HA VPN.

HA VPN is a high availability Cloud VPN solution that lets you securely connect your on-premises network to your Virtual Private Cloud (VPC) network through an IPsec VPN connection in a single region. HA VPN provides an SLA of 99.99% service availability. To guarantee a 99.99% availability SLA for HA VPN connections, you must properly configure two or four tunnels from your HA VPN gateway to your peer VPN gateway or to another HA VPN gateway.

When you create an HA VPN gateway, Google Cloud automatically chooses two external IP addresses, one for each of its fixed number of two interfaces. Each IP address is automatically chosen from a unique address pool to support high availability.

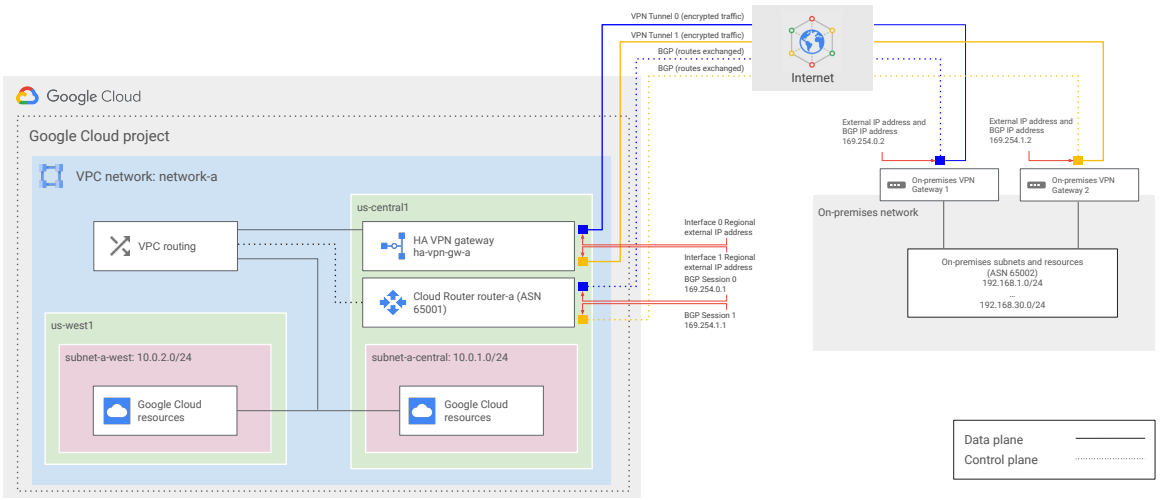
Each of the HA VPN gateway interfaces supports multiple tunnels. You can also create multiple HA VPN gateways. When you delete the HA VPN gateway, Google Cloud releases the IP addresses for reuse. You can configure an HA VPN gateway with only one active interface and one external IP address; however, this configuration does not provide a 99.99% service availability SLA. VPN tunnels connected to HA VPN gateways must use dynamic (BGP) routing. Depending on the way that you configure route priorities for HA VPN tunnels, you can create an active/active or active/passive routing configuration.

HA VPN supports site-to-site VPN in one of the following recommended topologies or configuration scenarios:

- An HA VPN gateway to peer VPN devices
- An HA VPN gateway to an Amazon Web Services (AWS) virtual private gateway
- Two HA VPN gateways connected to each other

Let's explore these configurations in a bit more detail.

HA VPN to peer VPN gateway topology



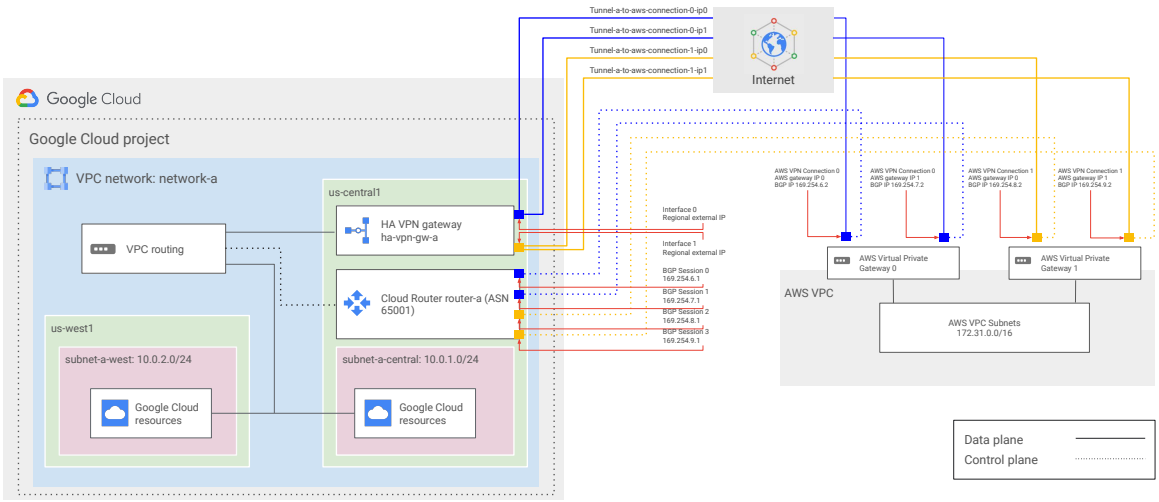
There are three typical peer gateway configurations for HA VPN. An HA VPN gateway to two separate peer VPN devices, each with its own IP address, an HA VPN gateway to one peer VPN device that uses two separate IP addresses and an HA VPN gateway to one peer VPN device that uses one IP address.

Let's walk through an example. In this topology, one HA VPN gateway connects to two peer devices. Each peer device has one interface and one external IP address. The HA VPN gateway uses two tunnels, one tunnel to each peer device. If your peer-side gateway is hardware-based, having a second peer-side gateway provides redundancy and failover on that side of the connection.

A second physical gateway lets you take one of the gateways offline for software upgrades or other scheduled maintenance. It also protects you if there is a failure in one of the devices.

In Google Cloud, the REDUNDANCY_TYPE for this configuration takes the value TWO_IPS_REDUNDANCY. The example shown here provides 99.99% availability.

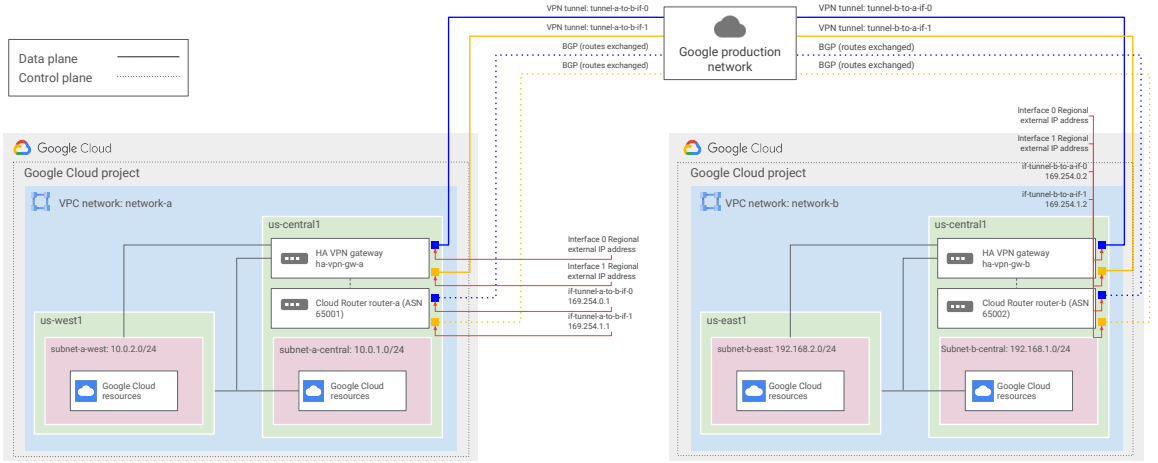
HA VPN to AWS peer gateway topology



When configuring an HA VPN external VPN gateway to Amazon Web Services (AWS), you can use either a transit gateway or a virtual private gateway. Only the transit gateway supports equal-cost multipath (ECMP) routing. When enabled, ECMP equally distributes traffic across active tunnels. Let's walk through an example.

In this topology, there are three major gateway components to set up for this configuration. An HA VPN gateway in Google Cloud with two interfaces, two AWS virtual private gateways, which connect to your HA VPN gateway, and an external VPN gateway resource in Google Cloud that represents your AWS virtual private gateway. This resource provides information to Google Cloud about your AWS gateway. The supported AWS configuration uses a total of four tunnels. Two tunnels from one AWS virtual private gateway to one interface of the HA VPN gateway, and two tunnels from the other AWS virtual private gateway to the other interface of the HA VPN gateway.

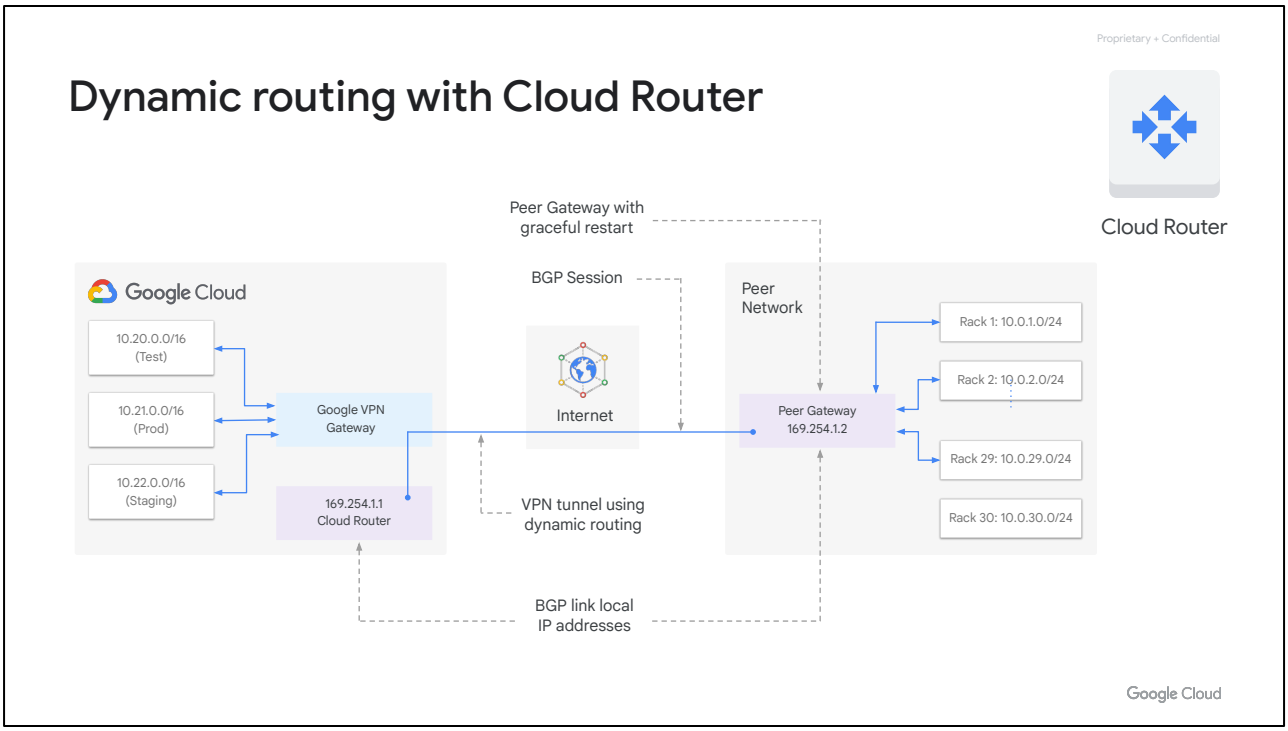
HA VPN between Google Cloud networks topology



You can connect two Google Cloud VPC networks together by using an HA VPN gateway in each network. The configuration shown provides 99.99% availability. From the perspective of each HA VPN gateway you create two tunnels. You connect `interface 0` on one HA VPN gateway to `interface 0` on the other HA VPN, and `interface 1` on one HA VPN gateway to `interface 1` on the other HA VPN.

For more information on HA VPN and moving to HA VPN, refer to the documentation links in the course resources.

[Cloud VPN topologies: https://cloud.google.com/network-connectivity/docs/vpn/concepts/topologies]
[Moving to HA VPN: https://cloud.google.com/network-connectivity/docs/vpn/how-to/moving-to-ha-vpn]



We mentioned earlier that Cloud VPN supports both static and dynamic routes. In order to use dynamic routes, you need to configure Cloud Routers. Cloud Router can manage routes for a Cloud VPN tunnel using Border Gateway Protocol, or BGP. This routing method allows for routes to be updated and exchanged without changing the tunnel configuration.

For example, this diagram shows two different regional subnets in a VPC network, namely Test and Prod. The on-premises network has 29 subnets, and the two networks are connected through Cloud VPN tunnels. Now, how would you handle adding new subnets? For example, how would you add a new “Staging” subnet in the Google Cloud network and a new on-premises 10.0.30.0/24 subnet to handle growing traffic in your data center?

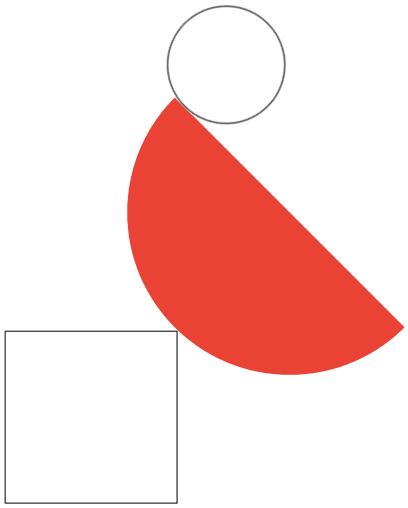
To automatically propagate network configuration changes, the VPN tunnel uses Cloud Router to establish a BGP session between the VPC and the on-premises VPN gateway, which must support BGP. The new subnets are then seamlessly advertised between networks. This means that instances in the new subnets can start sending and receiving traffic immediately, as you will explore in the upcoming lab.

To set up BGP, an additional IP address has to be assigned to each end of the VPN tunnel. These two IP addresses must be link-local IP addresses, belonging to the IP

address range 169.254.0.0/16. These addresses are not part of IP address space of either network and are used exclusively for establishing a BGP session.

Lab Intro

Configuring Google Cloud HA VPN



Let’s apply what we just covered.

In this lab you create a global VPC called vpc-demo, with two custom subnets in us-east1 and us-central1.

In this VPC, you add a Compute Engine instance in each region.

You then create a second VPC called on-prem to simulate a customer's on-premises data center. In this second VPC, you add a subnet in region us-central1 and a Compute Engine instance running in this region.

Finally, you add an HA VPN and a cloud router in each VPC and run two tunnels from each HA VPN gateway before testing the configuration to verify the 99.99% SLA.

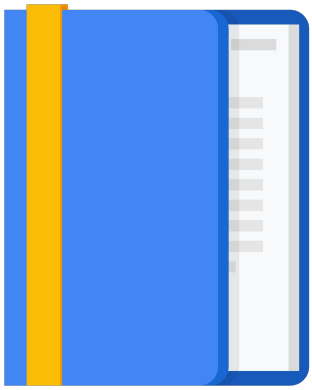
Agenda

Cloud VPN

Lab

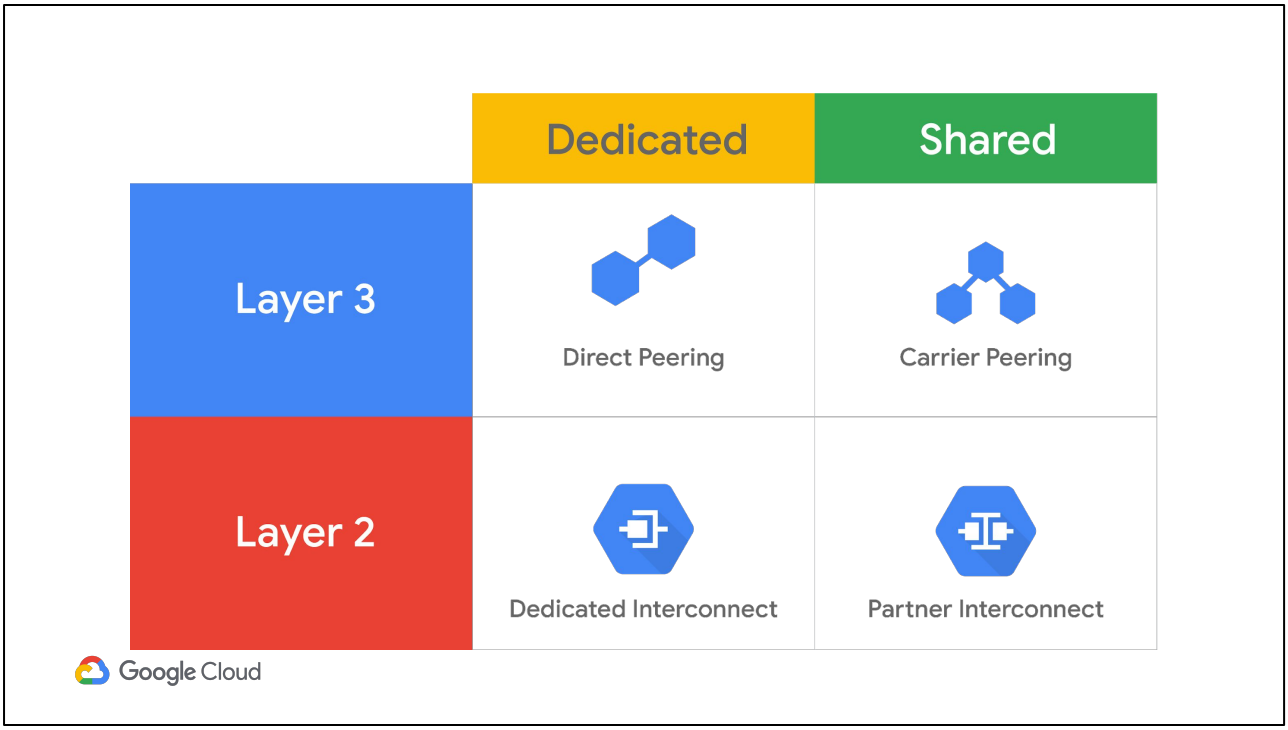
Cloud Interconnect and Peering

Sharing VPC Networks



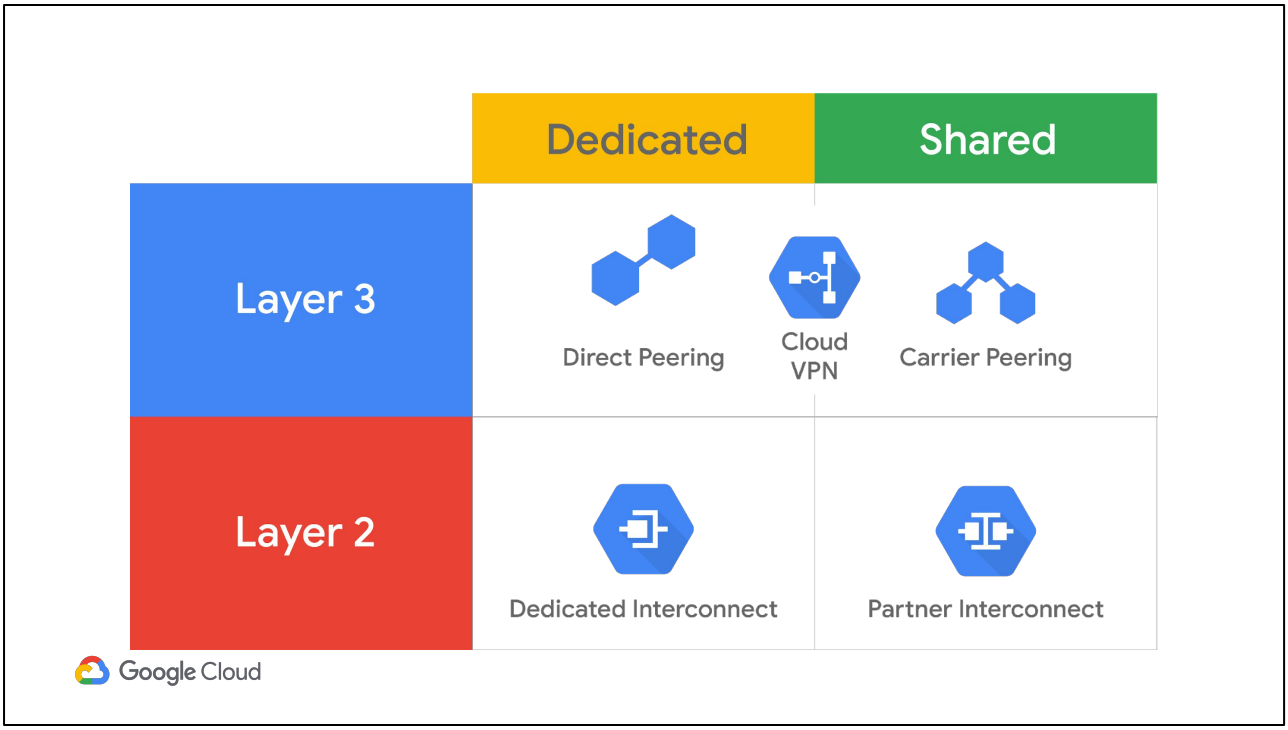
Google Cloud

Next, let's talk about the Cloud Interconnect and Peering services.



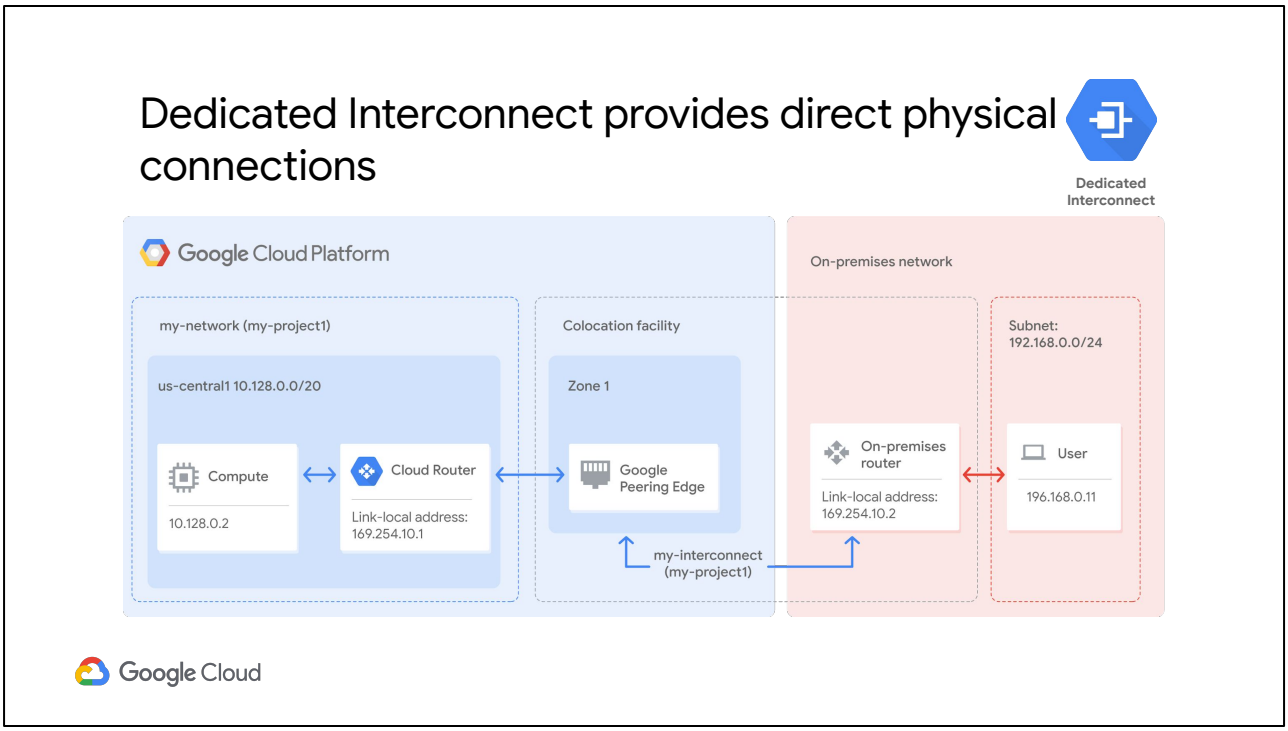
There are different Cloud Interconnect and Peering services available to connect your infrastructure to Google's network. These services can be split into Dedicated versus Shared connections and Layer 2 versus Layer 3 connections. The services are Direct Peering, Carrier Peering, Dedicated Interconnect, and Partner Interconnect.

Dedicated connections provide a direct connection to Google's network, but shared connections provide a connection to Google's network through a partner. Layer 2 connections use a VLAN that pipes directly into your GCP environment, providing connectivity to internal IP addresses in the RFC 1918 address space. Layer 3 connections provide access to Google Workspace services, YouTube, and Google Cloud APIs using public IP addresses.



Now, as I just explained earlier, Google also offers its own Virtual Private Network service, called Cloud VPN. This service uses the public internet, but traffic is encrypted and provides access to internal IP addresses. That's why Cloud VPN is a useful addition to Direct Peering and Carrier Peering.

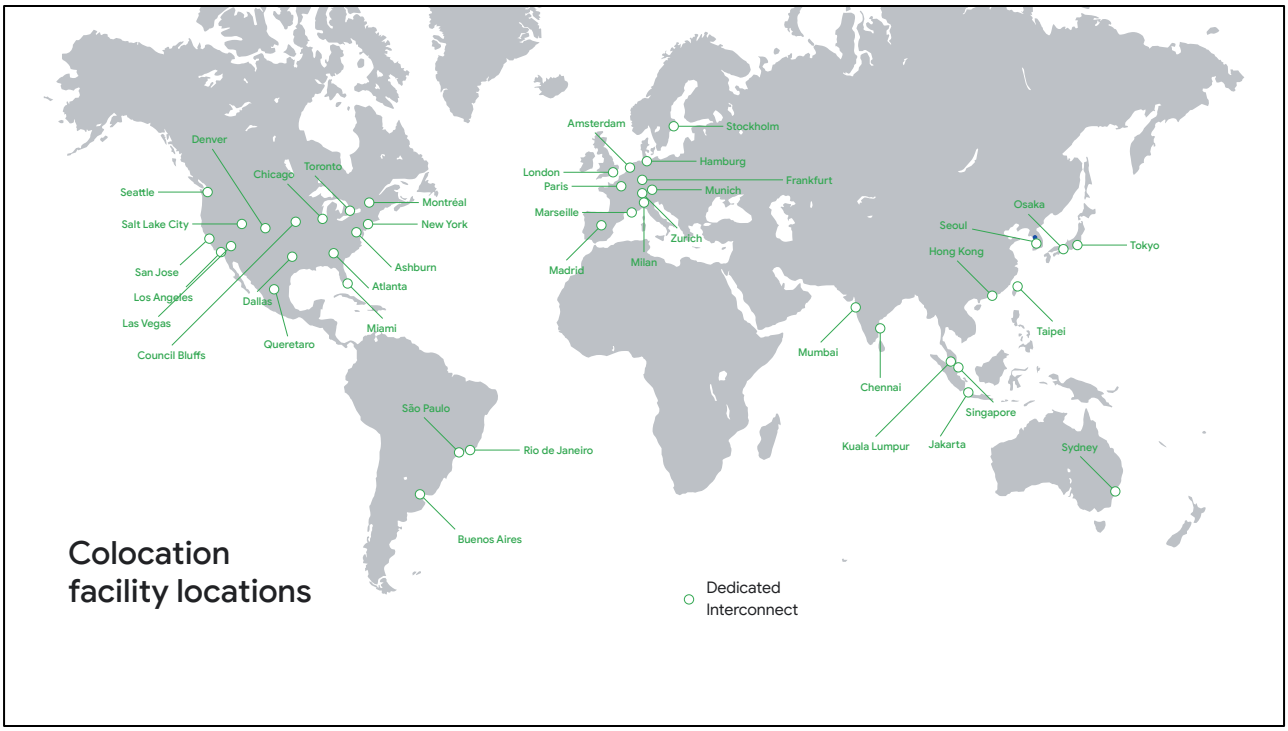
Let me explain the Cloud Interconnect and Peering services separately first and then I'll provide some guidance on choosing the right connection.



Dedicated Interconnect provides direct physical connections between your on-premises network and Google's network. This enables you to transfer large amounts of data between networks, which can be more cost-effective than purchasing additional bandwidth over the public internet.

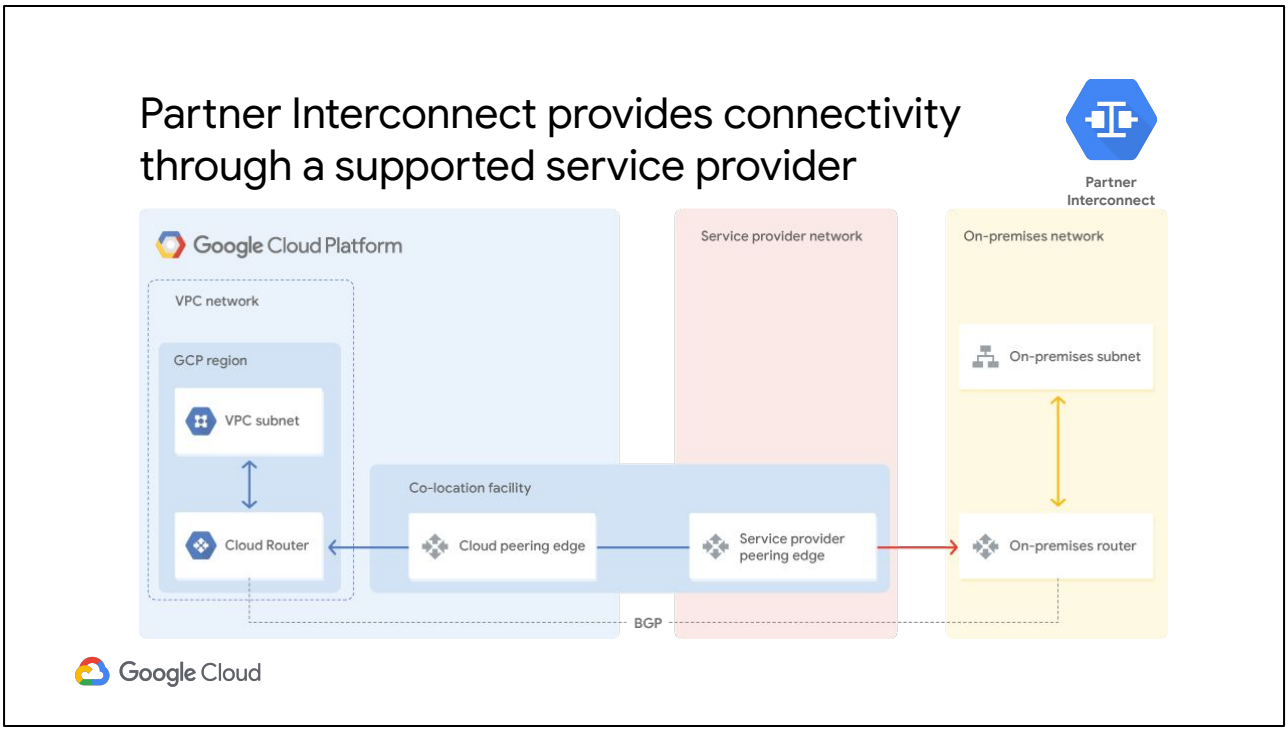
In order to use Dedicated Interconnect, you need to provision a cross connect between the Google network and your own router in a common colocation facility, as shown in this diagram. To exchange routes between the networks, you configure a BGP session over the interconnect between the Cloud Router and the on-premises router. This will allow user traffic from the on-premises network to reach GCP resources on the VPC network, and vice versa.

Dedicated Interconnect can be configured to offer a 99.9% or a 99.99% uptime SLA. See the Dedicated Interconnect documentation for details on how to achieve these SLAs [https://cloud.google.com/interconnect/docs/concepts/dedicated-overview#redundancy].



In order to use Dedicated Interconnect, your network must physically meet Google's network in a supported colocation facility. This map shows the locations where you can create dedicated connections. For a full list of these locations, see the links section of this video: [https://cloud.google.com/interconnect/docs/concepts/colocation-facilities].

Now, you might look at this map and say, "well I am nowhere near one of those locations." That's when you want to consider Partner Interconnect.



Partner Interconnect provides connectivity between your on-premises network and your VPC network through a supported service provider. This is useful if your data center is in a physical location that cannot reach a Dedicated Interconnect colocation facility or if your data needs don't warrant a Dedicated Interconnect.

In order to use Partner Interconnect, you work with a supported service provider to connect your VPC and on-premises networks. For a full list of providers, see the links section of this video:
[https://cloud.google.com/interconnect/docs/concepts/service-providers]

These service providers have existing physical connections to Google's network that they make available for their customers to use. After you establish connectivity with a service provider, you can request a Partner Interconnect connection from your service provider. Then, you establish a BGP session between your Cloud Router and on-premises router to start passing traffic between your networks via the service provider's network.

Partner Interconnect can be configured to offer a 99.9% or a 99.99% uptime SLA between Google and the service provider. See the Partner Interconnect documentation for details on how to achieve these SLAs [https://cloud.google.com/interconnect/docs/concepts/partner-overview#redundancy].

Comparison of Interconnect options

Connection	Provides	Capacity	Requirements	Access Type
IPsec VPN tunnel	Encrypted tunnel to VPC networks through the public internet	1.5-3 Gbps per tunnel	On-premises VPN gateway	Internal IP addresses
Dedicated Interconnect	Dedicated, direct connection to VPC networks	10 Gbps per link 100 Gbps BETA	Connection in colocation facility	
Partner Interconnect	Dedicated bandwidth, connection to VPC network through a service provider	50 Mbps – 10 Gbps per connection	Service provider	

Google Cloud

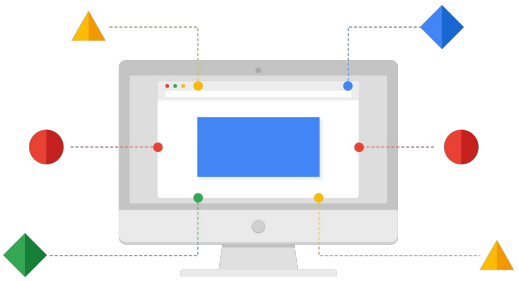
Let me compare the interconnect options that we just discussed. All of these options provide internal IP address access between resources in your on-premises network and in your VPC network. The main differences are the connection capacity and the requirements for using a service.

- The IPsec VPN tunnels that Cloud VPN offers have a capacity of 1.5 to 3 Gbps per tunnel and require a VPN device on your on-premises network. The 1.5-Gbps capacity applies to traffic that traverses the public internet, and the 3-Gbps capacity applies to traffic that is traversing a direct peering link. You can configure multiple tunnels if you want to scale this capacity.
- Dedicated Interconnect has a capacity of 10 Gbps per link and requires you to have a connection in a Google-supported colocation facility. You can have up to 8 links to achieve multiples of 10 Gbps, but 10 Gbps is the minimum capacity. As of this recording, there is a beta feature that provides 100 Gbps per link with a maximum of 2 links. Keep in mind that features that are in beta are not covered by any SLA or deprecation policy and might be subject to backward-incompatible changes.
- Partner Interconnect has a capacity of 50 Mbps to 10 Gbps per connection, and requirements depend on the service provider.

My recommendation is to start with VPN tunnels. When you need enterprise-grade connections to GCP, switch to Dedicated Interconnect or Partner Interconnect, depending on your proximity to a colocation facility and your capacity requirements.

Direct Peering provides a direct connection between your business network and Google's

- Broad-reaching edge network locations
- Exchange BGP routes
- Reach all of Google's services
- Peering requirements
- No SLA



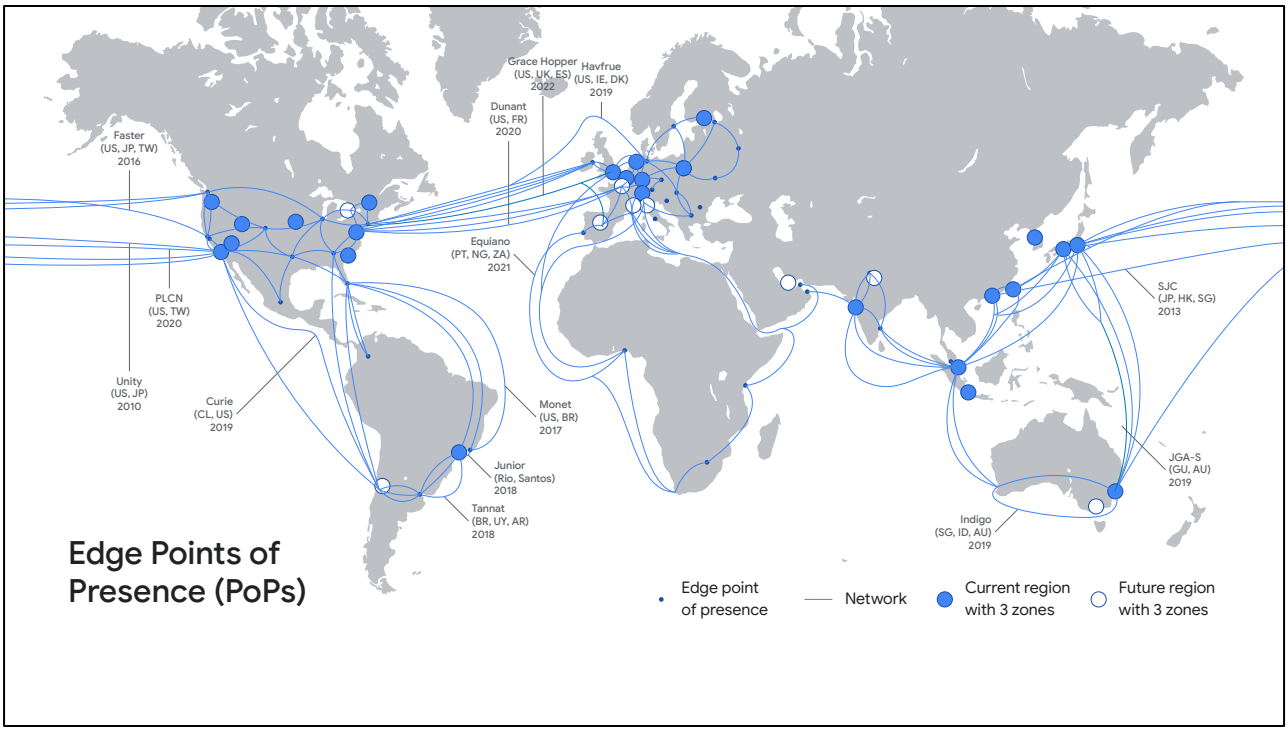
Google Cloud

Let's talk about the Cloud Peering services, which are Direct Peering and Carrier Peering. These services are useful when you require access to Google and Google Cloud properties.

Google allows you to establish a Direct Peering connection between your business network and Google's. With this connection you will be able to exchange internet traffic between your network and Google's at one of Google's broad-reaching edge network locations.

Direct Peering with Google is done by exchanging BGP routes between Google and the peering entity. After a Direct Peering connection is in place, you can use it to reach all of Google's services, including the full suite of Google Cloud Platform products. Unlike Dedicated Interconnect, Direct Peering does not have an SLA.

In order to use Direct Peering, you need to satisfy the peering requirements in the links section of this video [https://peering.google.com/#/options/peering].



GCP's Edge Points of Presence, or PoPs, are where Google's network connects to the rest of the internet via peering. PoPs are present on over 90 internet exchanges and at over 100 interconnection facilities around the world.

For more information about these exchange points and facilities, I recommend looking at Google's PeeringDB entries, which are linked below this video [https://www.peeringdb.com/asn/15169 and https://www.peeringdb.com/net/4319].

If you look at this map and say "Hey, I am nowhere near one of those locations," you will want to consider Carrier Peering.

Carrier Peering provides connectivity through a supported partner

- Carrier Peering partner
- Reach all of Google's services
- Partner requirements
- No SLA

Google Cloud

If you require access to Google public infrastructure and cannot satisfy Google's peering requirements, you can connect via a Carrier Peering partner. Work directly with your service provider to get the connection you need and to understand the partner's requirements. For a full list of available service providers, see the links section of this video [https://cloud.google.com/interconnect/docs/how-to/carrier-peering#service_providers] .

Now, just like Direct Peering, Carrier Peering does not have an SLA.

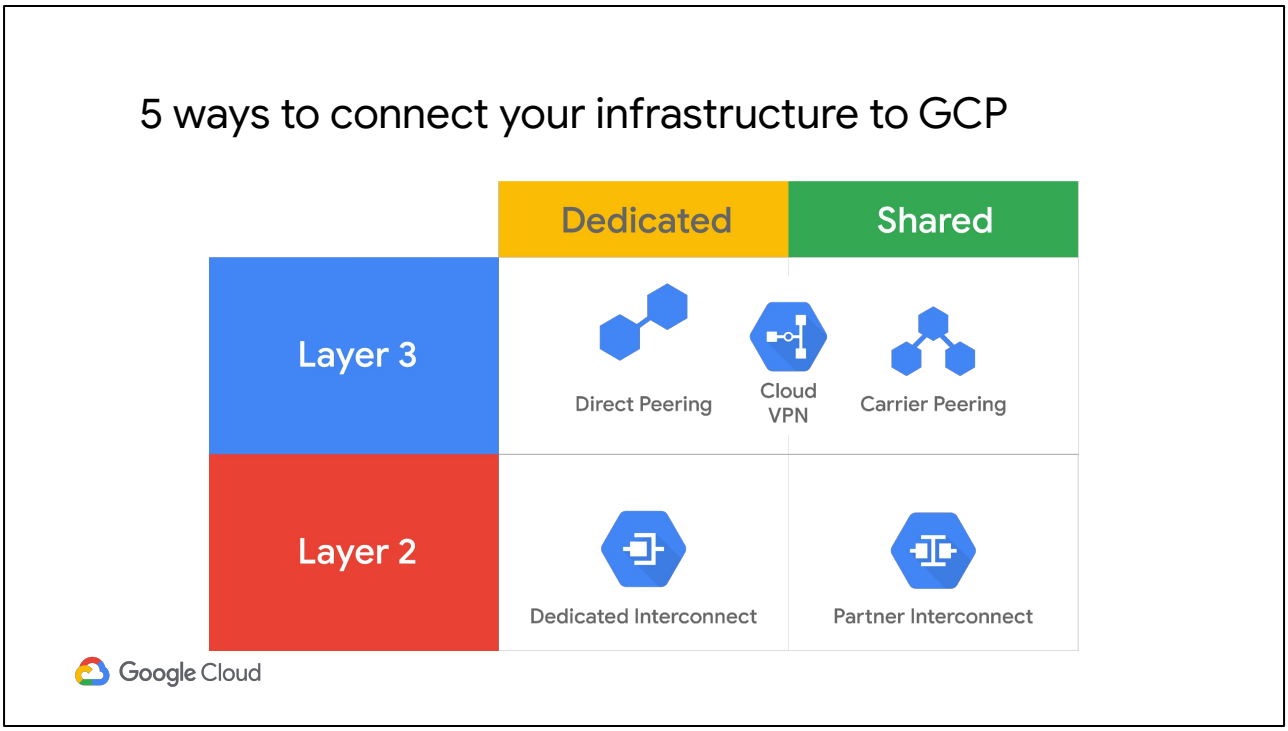
Comparison of Peering options

Connection	Provides	Capacity	Requirements	Access Type
Direct Peering	Dedicated, direct connection to Google's network	10 Gbps Per link	Connection in GCP PoPs	Public IP addresses
Carrier Peering	Peering through service provider to Google's public network	Varies based on partner offering	Service provider	

Google Cloud

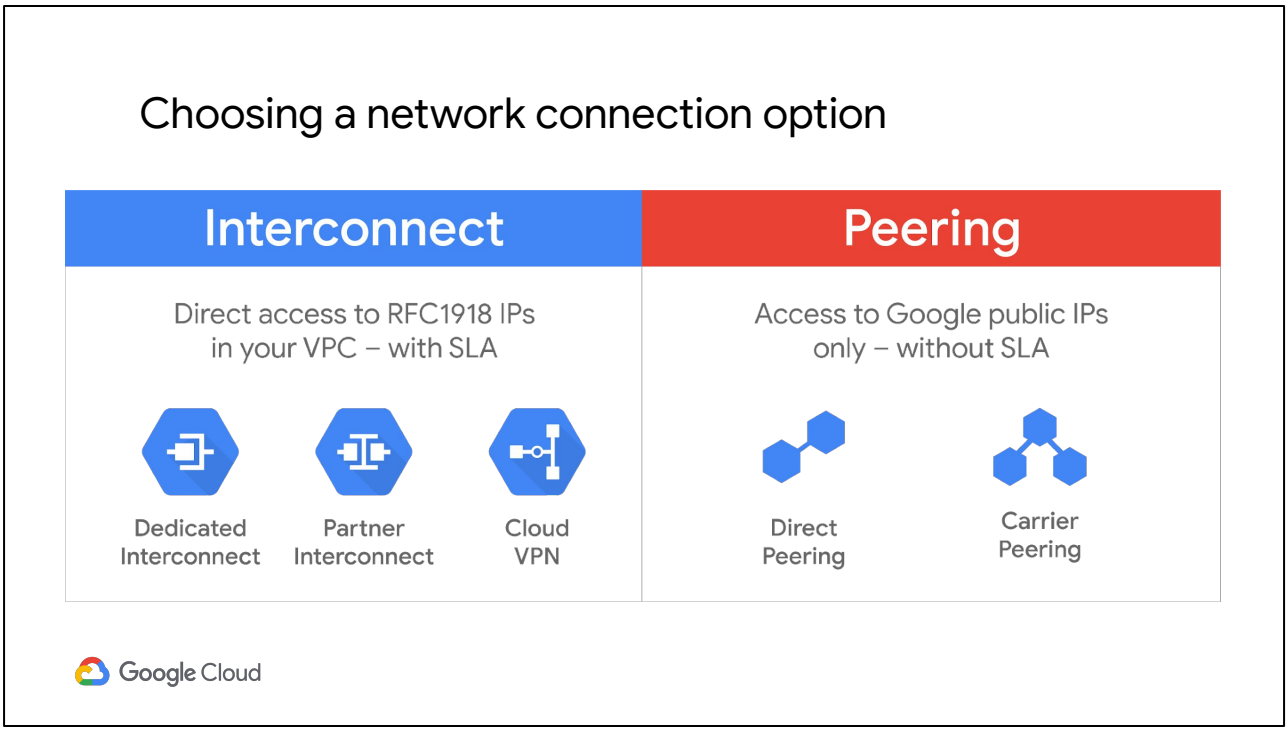
Let me compare the peering options that we just discussed. All of these options provide public IP address access to all of Google's services. The main differences are capacity and the requirements for using a service.

- Direct Peering has a capacity of 10 Gbps per link and requires you to have a connection in a GCP Edge Point of Presence.
- Carrier Peering's capacity and requirements vary depending on the service provider that you work with.



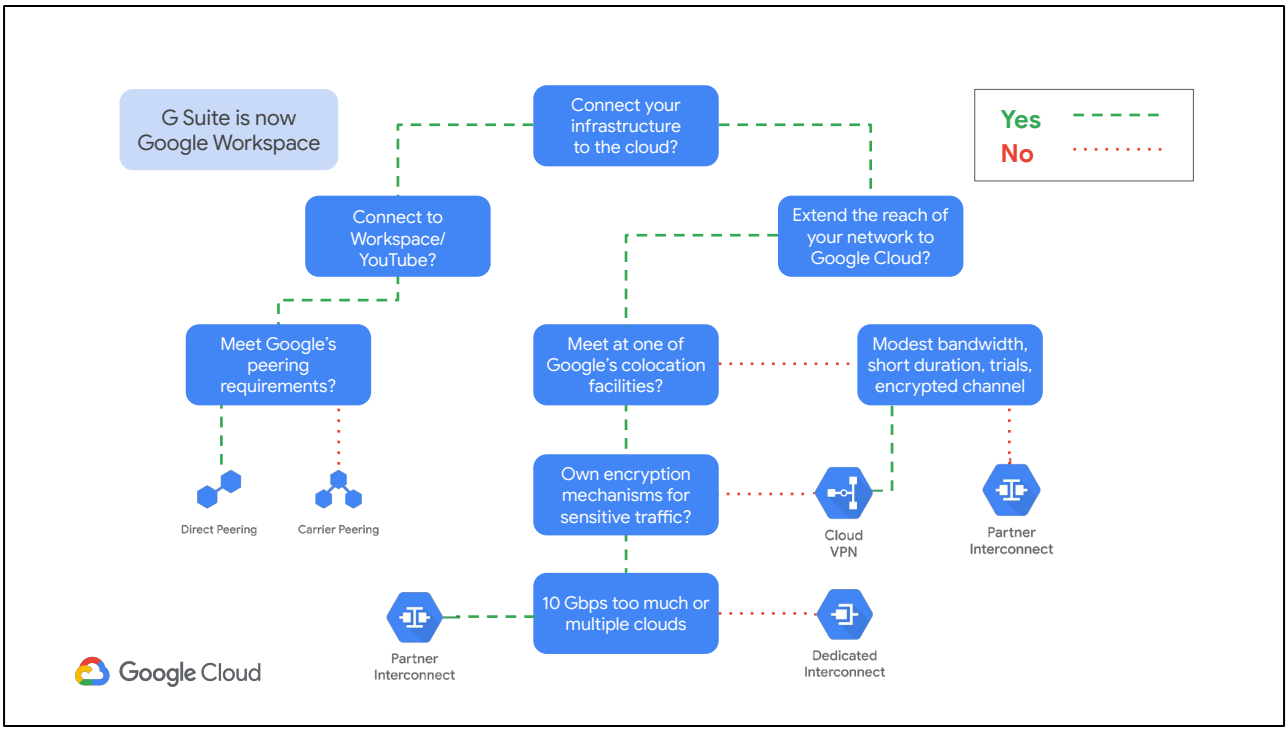
Now that we have discussed all the different connection services, let me help you determine which service best meets your hybrid connectivity needs.

I started this lesson by introducing the 5 different ways to connect your infrastructure to GCP. I split these services into Dedicated versus Shared connections and Layer 2 versus Layer 3 connections.



Another way to organize these services is by Interconnect services and by Peering services.

Interconnect services provide direct access to RFC1918 IP addresses in your VPC, with an SLA. Peering services, in contrast, offer access to Google public IP addresses only, without an SLA.



Another way to choose the right service that meets your needs is with a flow diagram. Let me walk you through this diagram from the top, using the assumption that you want to extend your infrastructure to the cloud.

Ask yourself whether you need to extend your network for Workspace services, YouTube, or Google Cloud APIs. If you do, choose one of the Peering services. If you can meet Google's Direct Peering requirements, choose Direct Peering; otherwise, choose Carrier Peering.

If you don't need to extend your network for Workspace services or Google Cloud APIs but want to extend the reach of your network to Google Cloud, you want to pick one of the Interconnect services. If you cannot meet Google at one of its colocation facilities, choose Cloud VPN or Partner Interconnect. This choice will depend on your bandwidth and encryption requirements, along with the purpose of the connection. Specifically, if you have modest bandwidth needs, will use the connection for short durations and trials, and require an encrypted channel, choose Cloud VPN; otherwise, choose Partner Interconnect.

If you can meet Google at one of its colocation facilities, you might jump to Dedicated Interconnect; however, if you cannot provide your own encryption mechanisms for sensitive traffic, feel that a 10-Gbps connection is too big, or want access to multiple clouds, you'll want to consider Cloud VPN or Partner Interconnect instead.

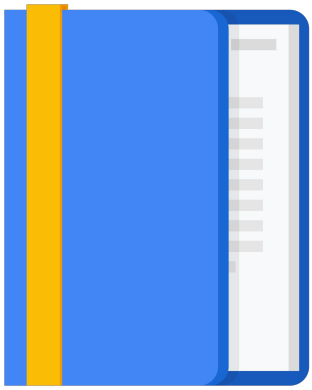
Agenda

Cloud VPN

Lab

Cloud Interconnect and Peering

Sharing VPC Networks

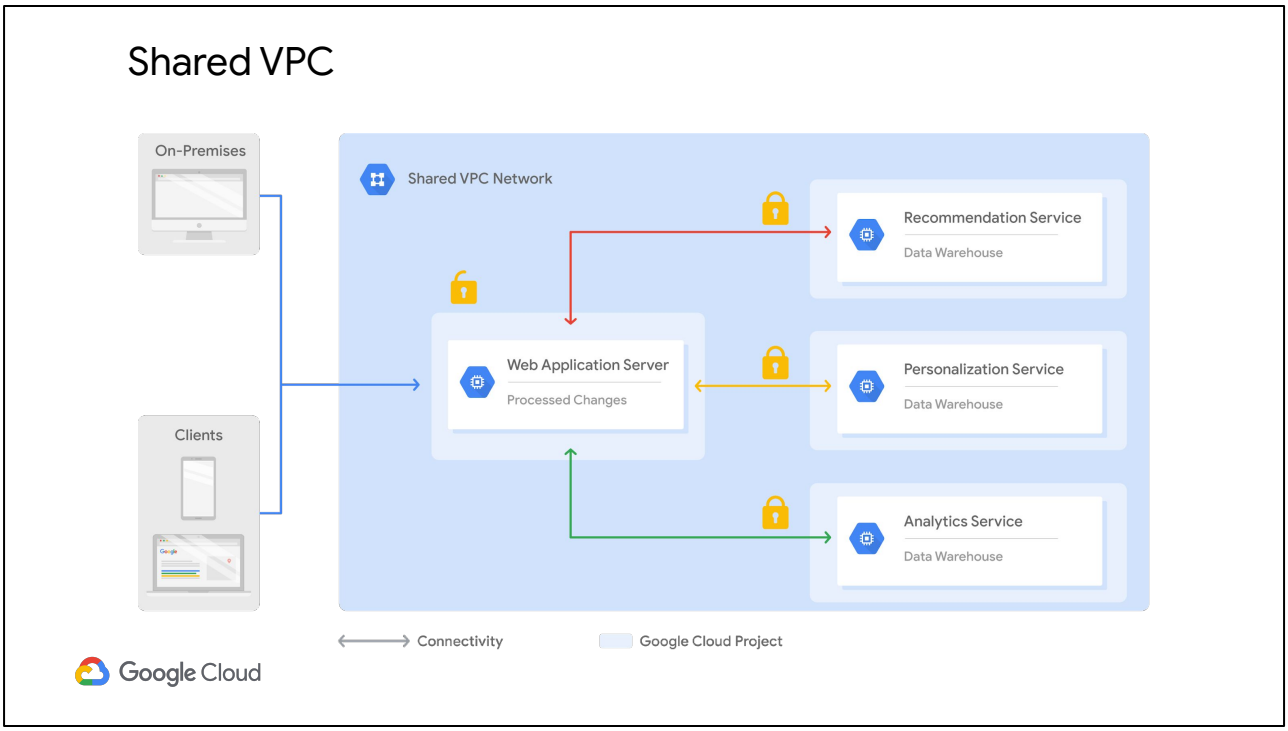


Google Cloud

Let’s move our attention from hybrid connectivity to sharing VPC networks.

In the simplest cloud environment, a single project might have one VPC network, spanning many regions, with VM instances hosting very large and complicated applications. However, many organizations commonly deploy multiple, isolated projects with multiple VPC networks and subnets.

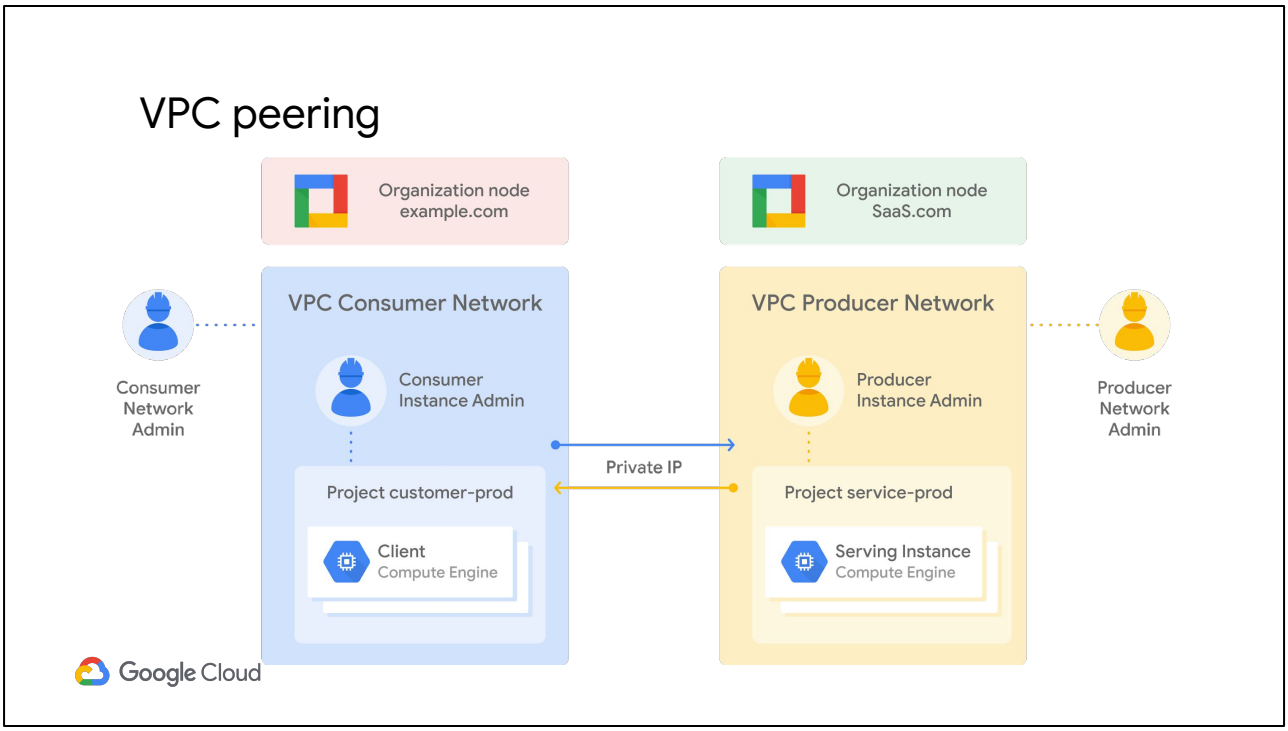
In this lesson, we are going to cover two configurations for sharing VPC networks across GCP projects. First, we will go over shared VPC, which allows you to share a network across several projects in your GCP organization. Then, we will go over VPC Network Peering, which allows you to configure private communication across projects in the same or different organizations.



Shared VPC allows an organization to connect resources from multiple projects to a common VPC network. This allows the resources to communicate with each other securely and efficiently using internal IPs from that network.

For example, in this diagram there is one network that belongs to the Web Application Server's project. This network is shared with three other projects, namely the Recommendation Service, the Personalization Service, and the Analytics Service. Each of those service projects has instances that are in the same network as the Web Application Server and allow for private communication to that server, using internal IP addresses. The Web Application Server communicates with clients and on-premises using the server's external IP address. The backend services, in contrast, cannot be reached externally because they only communicate using internal IP addresses.

When you use shared VPC, you designate a project as a host project and attach one or more other service projects to it. In this case, the Web Application Server's project is the host project, and the three other projects are the service projects. The overall VPC network is called the shared VPC network.



VPC Network Peering, in contrast, allows private RFC 1918 connectivity across two VPC networks, regardless of whether they belong to the same project or the same organization. Now, remember that each VPC network will have firewall rules that define what traffic is allowed or denied between the networks.

For example, in this diagram there are two organizations that represent a consumer and a producer, respectively. Each organization has its own organization node, VPC network, VM instances, Network Admin, and Instance Admin. In order for VPC Network Peering to be established successfully, the Producer Network Admin needs to peer the Producer Network with the Consumer Network, and the Consumer Network Admin needs to peer the Consumer Network with the Producer Network. When both peering connections are created, the VPC Network Peering session becomes Active and routes are exchanged. This allows the virtual machine instances to communicate privately using their internal IP addresses.

VPC Network Peering is a decentralized or distributed approach to multi-project networking, because each VPC network may remain under the control of separate administrator groups and maintains its own global firewall and routing tables. Historically, such projects would consider external IP addresses or VPNs to facilitate private communication between VPC networks. However, VPC Network Peering does not incur the network latency, security, and cost drawbacks that are present when using external IP addresses or VPNs.

Shared VPC vs. VPC peering

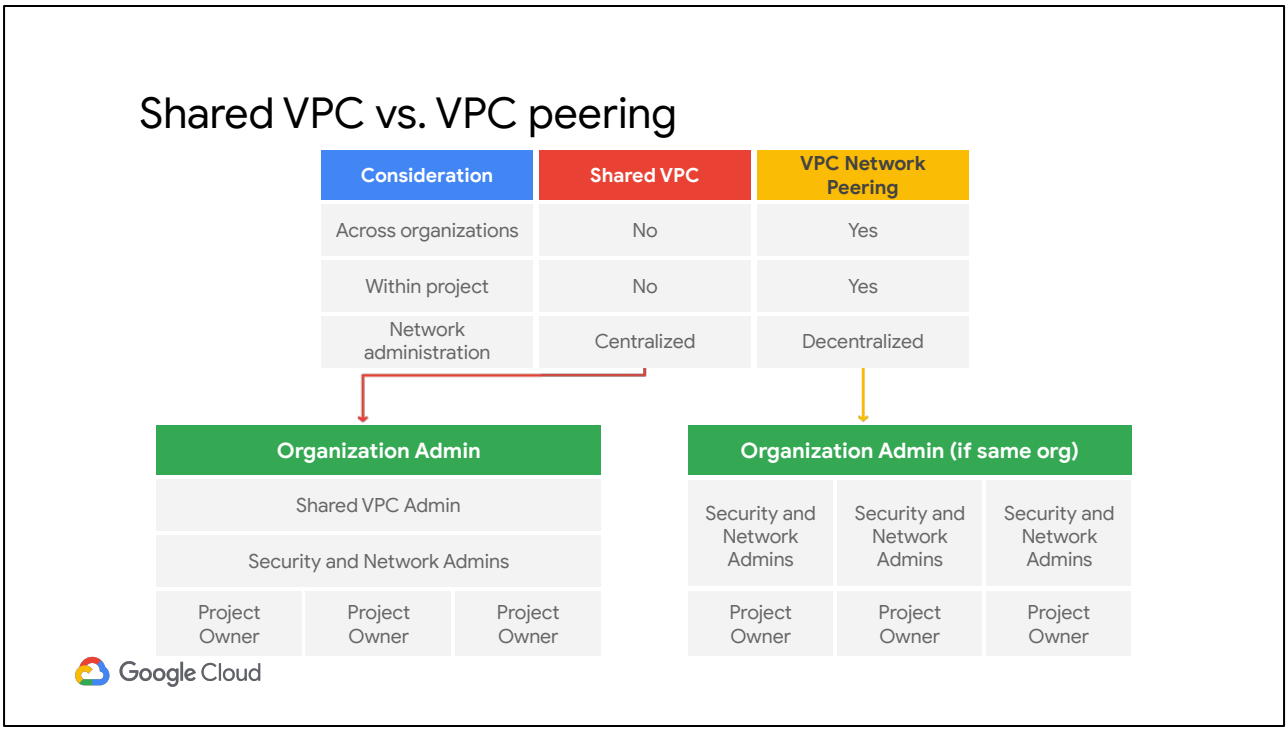
Consideration	Shared VPC	VPC Network Peering
Across organizations	No	Yes
Within project	No	Yes
Network administration	Centralized	Decentralized

Google Cloud

Now that we've talked about Shared VPC and VPC Network Peering, let me compare both of these configurations to help you decide which is appropriate for a given situation.

If you want to configure private communication between VPC networks in different organizations, you have to use VPC Network Peering. Shared VPC only works within the same organization.

Somewhat similarly, if you want to configure private communication between VPC networks in the same project, you have to use VPC Network Peering. This doesn't mean that the networks need to be in the same project, but they can be. Shared VPC only works across projects.



In my opinion, the biggest difference between the two configurations is the network administration models. Shared VPC is a centralized approach to multi-project networking, because security and network policy occurs in a single designated VPC network. In contrast, VPC Network Peering is a decentralized approach, because each VPC network can remain under the control of separate administrator groups and maintains its own global firewall and routing tables.

Review

Interconnecting Networks

In this module, we looked at the five different ways of connecting your infrastructure to GCP, which are Dedicated Interconnect, Partner Interconnect, Cloud VPN, Direct Peering, and Carrier Peering. I also gave you some guidance on how to choose between different services. Remember, you might start out using one service, and as your requirements change or new colocation facilities open, switch to a different service.

I also gave you a brief overview of shared VPC and VPC Network Peering, which are two configurations for sharing VPC networks across GCP projects.