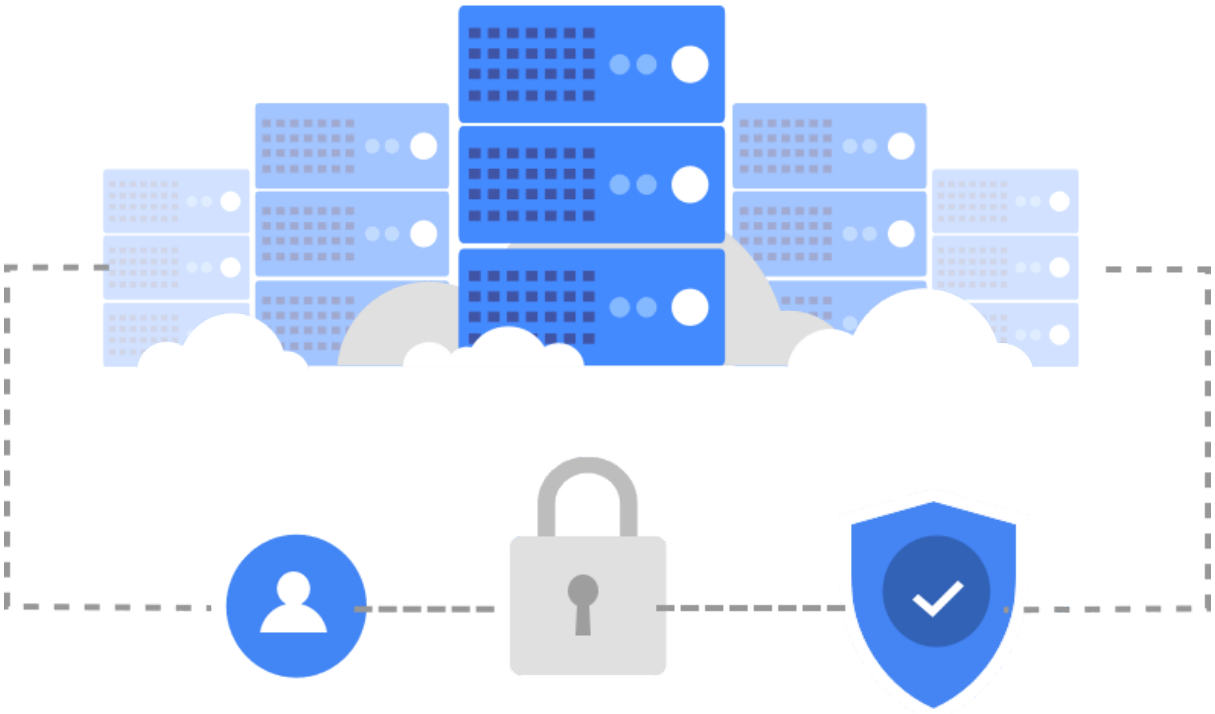
Deploying on Google Cloud Platform with HSCN - Reference Architecture



January 2021

Reference Architecture

Overview

Health and social care network (HSCN) is intended for health and social care organisations to access and exchange digital information in England, UK.

Organisations that provide health and social care services will deploy solutions using HSCN if they need to access systems already available on HSCN such as national SPINE and Personal Demographic Service (PDS) services. Any solution that involves collaboration by more than one organisation needs to be deployed using HSCN. This includes Integrated Care Systems (ICS).

Organisations do not need to use HSCN if the solution is only used by users of that organisation and the solution does not integrate with any systems deployed on HSCN.

In all cases customers can choose to deploy their solution using the public cloud such as Google's Cloud Platform (GCP). The focus of this paper is to look at how you can deploy solutions on GCP and connect to HSCN using best practices. This paper is intended for Trusts and 3rd parties who want to deploy HSCN connected solutions on GCP.

Compliance

Customers of HSCN are responsible for signing HSCN Connection Agreements to get access to the network.

They must also complete a Data Security and Protection Toolkit (DSP Toolkit) to provide assurances that they are practicing good data security in order to gain access to the services that are available on HSCN.

Google LLC is an organisation external to health and social care. It contracts with customers of HSCN to provide cloud services acting as a data processor. DSP Toolkit categorises Google as a ‘company’. Google has completed its own DSP Toolkit assessment as a company. You can manually search for the Google compliance status using the DSP Toolkit portal. Use the organisation name ‘Google LLC’ or the organisation code ‘8JE14’.

The best practice and patterns provided in this publication are intended to act as a guide only. Customers and their delivery partners are responsible for ensuring that any deployed solutions meet the requirements including the compliance standards.

Requirements

The following section outlines the key requirements for any health and social care solution that uses HSCN.

Principles

Minimise risk to HSCN is something that is achieved through good security patterns and practices.

There must be no unknown paths between HSCN and the public Internet. A solution deployed on the public cloud can have paths that lead from and to the public internet. What is important is ensuring that these are known paths that have been securely designed and implemented based on the solution needs.

Assumptions

Solutions in a production environment will typically have requirements including high availability, low latency, operations management and maintainability.

These requirements will vary from solution to solution depending on its specific needs. It is important to take these into consideration when designing your solution on GCP and leveraging the appropriate services in the design.

Constraints

The solution must be compliant with the DSP Toolkit. Whilst Google Cloud is compliant with the DSP Toolkit, any organisation deploying on GCP must complete its own DSP Toolkit assessment.

The customer must also sign a HSCN Connection Agreement with a consumer network service provider (CNSP).

Architecture design for security

This section provides the guidance to help you design your solutions to meet your needs and obligations.

Foundation patterns

Google publishes best practices guides on its security best practices centre. These resources are informed guidance on helping secure GCP deployments and describe recommended configurations, architectures, suggested settings and other operational advice to design your solutions.

Google Cloud security foundations guide (Foundations) can be used as a template for any solution. It provides customers with curated, opinionated guidance together with accompanying automation using Terraform that helps you build a secure starting point for your GCP deployment.

The guidance is based on certain architectural decisions documented in Foundations, section 3 Google Cloud foundation design. If your assumptions are different then you need to make changes to the architecture used in the guidance.

HSCN Connectivity

GCP solution on HSCN

The architecture depicted in Figure 1 below shows how a solution deployed in GCP can connect with HSCN. The design highlights five potential key paths for any HSCN facing service deployed in GCP.

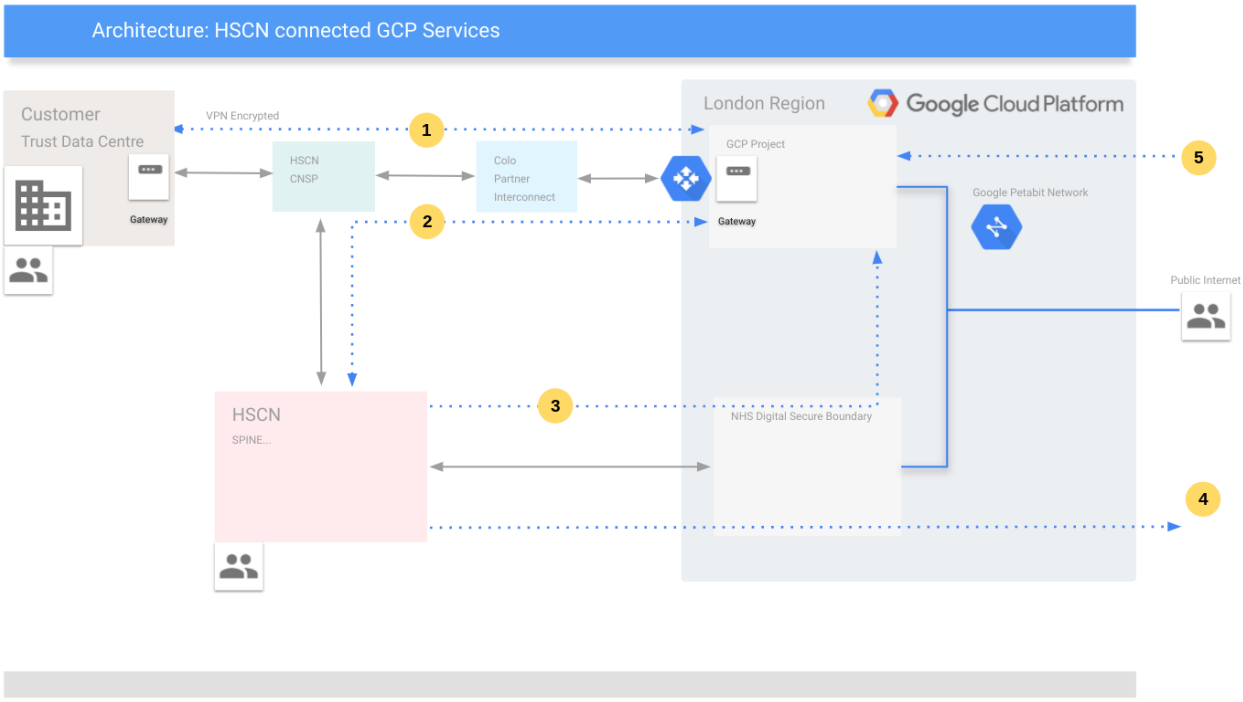


Figure 1: Architecture: HSCN connected GCP Services

Path 1 Customer site connectivity to GCP

Path 1 connects a customer data centre with GCP to send and receive data. It is used to support administration and depending on the solution it can also be used to send large volumes of data to GCP. This includes batch data and data streams such as HL7 v2 ADT messages from a Trust Integration Engine. As an example in GCP a solution may process data using Big Data and AI services. Any recommendations can be sent back to systems on-premises to support care or operations. The solution can use any of the GCP services to support its needs such, i.e. synchronise data for any disaster recovery solution.

The design is using GCP Partner Interconnect to provide a private link through a service provider to GCP. As of September 2020, it is not currently possible to use Cloud VPN with Interconnect. A 3rd party VPN solution such as StrongSwan can be used to provide end to end encryption over the Interconnect. This connectivity provides low latency and high throughput access required of typical solutions through a private link.

Google recommends a High Availability Interconnect deployment for production systems. You must work with service providers to establish connectivity between the customer network and the service providers network.

Depending on the requirements, instead of using Cloud Interconnect and 3rd party VPN, you could use Cloud VPN on top of public internet as an alternative to provide connectivity between the customer data centre and GCP. Refer to the HA Cloud VPN documentation for more details.

A solution may require data to be transferred to GCP, which requires careful consideration. Google provides best practice documentation for data transfer from assessing, planning, deployment and optimisation. You will need to decide whether you use an online or offline data transfer process, that is, use a network such as Partner Interconnect or a storage device. You can use the transfer calculator to understand how much time a transfer might take given the amount of data and bandwidth available. The data transfer options include gsutil command line and storage transfer service for online transfers, or transfer appliance for offline transfers.

The customer must use a HSCN connection supplier (CNSP) to connect the solution to the HSCN. Refer to the NHS Digital HSCN site for details of service providers. Some suppliers provide both HSCN and Interconnect services. See the network connectivity stewardship section for discussion of when you must consider using the same supplier for both connectivity services.

Path 2 Accessing HSCN

NHS Digital provides an IP address management service (IPAM) for HSCN. They will allocate an IP CIDR range for your HSCN solution as per their policy, e.g. for illustration only, 10.0.0.0/24 from the RFC 1918 private address space. We will refer to this as the HSCN-GCP subnet in Google Cloud.

This subnet is configured in GCP and any HSCN facing services are deployed into the network. Services that are not directly connected to HSCN can be segmented using a separate subnet and GCP Project, see Segmenting HSCN facing services.

Path 2 can be used by HSCN facing services to integrate with other services on HSCN, i.e. SPINE. The solution may perform a patient demographic search to confirm a patient's identity. It can also be used to allow systems deployed on HSCN to invoke your services deployed on GCP. HSCN provides its own DNS service, making it easier for others to locate your services.

Path 3 Accessing a GCP solution using Secure Boundary

NHS Digital provides a perimeter security solution for HSCN called Secure Boundary (SB). All traffic intended for the public internet is first sent through the SB service before it is routed to the internet.

SB is deployed in the GCP UK Region. This provides an advantage if a system deployed on HSCN needs to communicate with a system deployed on GCP using a secure public endpoint. The request will be routed through the SB solution. Once it is routed to the Internet, the request will be forwarded to your GCP solution without traversing the Internet as both SB and your solution are in the same GCP region on Google's network. Traffic between public IPs on Google ASN15169 always stays on Google's backbone network whether it's intra-region or inter-region.

Path 4 Accessing the Public Internet from HSCN

Any traffic destined for the public internet from HSCN will be routed through the SB service.

Path 5 Accessing the GCP solution from the Public Internet

Users can access the GCP solution from the public internet. Any solution is expected to follow good security practices.

Segmenting HSCN facing services

Figure 2 shows how any HSCN facing services can be segmented. Principles of least privilege and defense in depth are used. The best practice Foundations guidance provides fuller details of how you can secure your GCP deployments. This section summarises key patterns.

Network isolation

Consider designing the solution so that you only deploy HSCN facing services in a restricted HSCN connected network (HSCN-GCP subnet). Non HSCN services can be deployed to other networks as required. This further mitigates the risk to HSCN.

In GCP a Project is used to organise all your resources. It is also a billing boundary.

Shared VPC is used as the primary networking topology to enable centralised management of firewall rules and connectivity. A Shared VPC host project can be configured with one or more Shared VPC networks, which can then be leveraged by attached service projects. IAM is used to ensure specific administrators have access as part of an overall enterprise resource hierarchy.

Resources in Shared VPC networks can communicate with each other securely and efficiently across project boundaries using internal IP addresses. You can manage shared network resources such as subnets, routes, and firewalls from a central host project, so you can enforce consistent network policies across the projects.

VPC Service Controls provides perimeter protection for services that store highly sensitive data providing service level segmentation. This helps to ensure sensitive data can only be accessed from authorised networks. Only restricted access to allowed IP addresses and identities is permitted. You can control which Google Cloud services are accessible from the VPC network. To enable projects to communicate with projects across perimeters you need to create a service perimeter bridge.

Projects are created under Folders for each type of environment i.e. production and non-production. A Shared VPC is used for any services that do not require VPC Service Controls. A Restricted Shared VPC is used for HSCN-GCP subnet requiring VPC Service Controls.

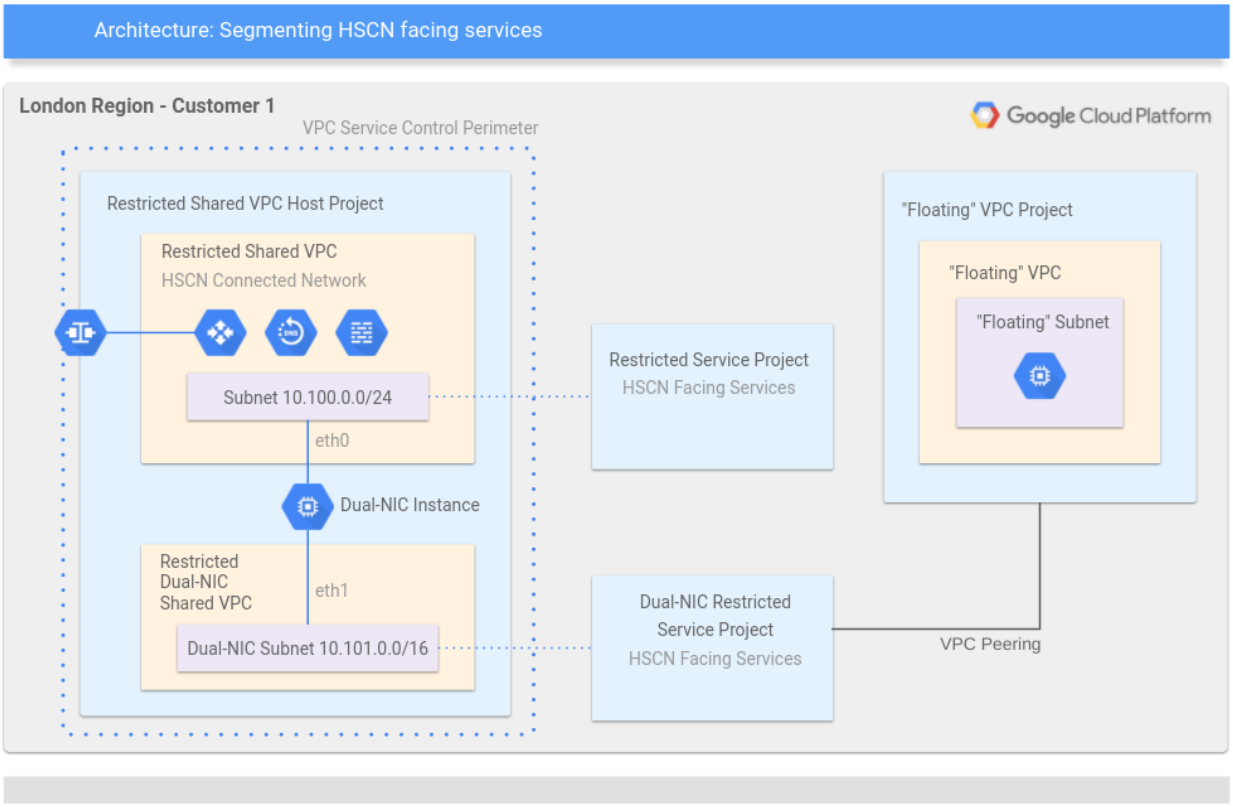


Figure 2: Architecture: Segmenting HSCN facing Services

Segmenting HSCN and non-HSCN services

Floating Projects have no direct connectivity to HSCN. They require access solely to GCP resources, however, in some cases they may use services that are connected to HSCN. These Projects can be modified to use a Shared VPC to provide centralised network management. They might be used for segmenting workloads such as batch or front end from the HSCN facing services.

You can use VPC Peering to enable internal IP address communication between VPC networks where required. Traffic stays in the Google network and does not traverse the public internet. Only directly peered networks can communicate where routes and firewall rules allow - transitive peering is not supported.

Managing network address ranges with Dual-NIC

A Dual-NIC pattern is used when connectivity to HSCN is required, but the IP address space requirements of the HSCN facing services is larger than the HSCN-GCP subnet CIDR range allocated by NHS Digital to the customer. A restricted Dual-NIC Shared VPC network is connected to HSCN-GCP subnet through a Dual-NIC enabled Compute Engine instance that

performs NAT. This allows resources in the restricted Dual-NIC Shared VPC network to communicate with resources in HSCN-GCP subnet through a NAT or a proxy service running on the Dual-NIC instance.

Cross Project multi NIC configurations require eth0 of the Dual-NIC instance to be connected to the Restricted Shared VPC network, and eth1 to be connected to a separate Dual-NIC Shared VPC. Both the Restricted Shared VPC and Dual-NIC VPC are in the same Restricted Shared VPC Host Project. Figure 2 shows how this approach expands the range of addresses that can access HSCN. The HSCN-GCP subnet address space is 10.100.0.0/24 or 252 hosts. The Dual-NIC subnet address space is 10.101.0.0/16 or 65,532 hosts (every GCP network reserves 4 IP addresses).

Mapping Kubernetes clusters to HSCN network ranges

Depending on the capacity and other requirements of a solution, you can deploy microservices directly in HSCN-GCP subnet using Kubernetes. In this case, the node IP range used for the Kubernetes Services should use the HSCN-GCP CIDR address range. Connectivity to the Kubernetes cluster services is provided through the Kubernetes node IP range. For example, you can expose a cluster service as an Internal load balancer with a virtual IP (VIP) taken from the node IP range. When traffic reaches a Kubernetes node through the VIP, the node forwards the traffic to the relevant cluster service and pods. You can read more about networking outside the cluster and private VPC native cluster IP ranges. You can also restrict the source ranges that can communicate with your service through the load balancer.

Managing Kubernetes egress to external destinations

For egress traffic originating from within the cluster to external destinations, you can use IP masquerading to ensure pod IP ranges are masqueraded behind the Kubernetes node IP. IP masquerading simply uses NAT to translate the pod IP ranges to the Kubernetes node IP. This ensures that all Kubernetes cluster traffic to external destinations come from the HSCN address range allocated. This gives great flexibility because the service and pod IP ranges do not have to be unique within the HSCN network; only the node IP ranges have to be unique.

Managing egress to the Internet

Cloud NAT can be used for any private Compute Instances, or for nodes and pods in a private kubernetes cluster to communicate with the Internet. Cloud NAT cannot be used to egress to HSCN. This negates the need for any instances to have public IP addresses further mitigating security risks.

Service accounts are a special kind of account used by applications or virtual machine instances to make authorised API calls to resources. The best practice Foundations provides general security guidance for service accounts using the principle of least privilege. Ensure service accounts are tighty scoped with permissions based on their use case.

Service accounts can be used to restrict access to Cloud NAT as an additional control.

Egress to the internet may be used by parts of the system that are not deployed in a Restricted Shared VPC (HSCN facing).

Network connectivity stewardship

There are two use cases that are important to consider with HSCN connectivity, the first is where a Trust deploys a solution to GCP for use by the Trust and their partners across health and social care. The second is where a 3rd party provides a solution (e.g. a SaaS application) to health and social care. Depending on the scenario, the entity that provisions and manages the partner network connectivity services is different. In the case of a 3rd party it can enable them to provide their service across health and social care using the same connectivity infrastructure.

Use case: Trust deploys solution in GCP

If a Trust owns the solution being deployed in GCP. They will be responsible for signing a HSCN Connection Agreement with NHS Digital. The Trust will also be responsible for provisioning services from a HSCN CNSP and a GCP Partner Interconnect service provider to support their solution network connectivity needs. In this scenario the HSCN and Partner Interconnect service providers can be the same or different.

Use case: A 3rd party service provider deploys a solution in GCP for health and social care entities

A 3rd party service provider may deploy a solution in GCP intended for use by different kinds of health and social care organisations. This may be enabled for specific organisations depending on commercial terms. Here, the 3rd party is responsible for signing the HSCN Connection Agreement with NHS Digital. The 3rd party will also be responsible for provisioning HSCN CNSP and a GCP Partner Interconnect service provider to support their solution network connectivity needs.

In this scenario the 3rd party owns and manages the environments in GCP as well as the network connectivity into HSCN. They provide services to customers across the system using the same HSCN connection. To enable this the 3rd party must use the same service provider for both HSCN and GCP connectivity. The service provider will be required to bridge connectivity between HSCN and GCP.

Once it is setup the 3rd party can request the HSCN CNSP to enable routes and firewall rules between its service on GCP and specific customers across health and social care that have signed up to use its service. This assumes that existing organisations have a connectivity to HSCN from their premises.

To create a secure link between the customer and the service in GCP the 3rd party will need to deploy a VPN gateway for each customer of their solution. A VPN gateway is deployed in a Restricted VPC network that uses HSCN CIDR range for all customers, lets call this HSCN-Network. This network is responsible for HSCN connectivity.

The VPN gateway will be deployed on a GCE instance that uses the Dual-NIC pattern. eth0 will be connected to HSCN-Network. eth1 is then connected to a Restricted Shared VPC network that uses a private CIDR range, let's call this Customer# Restricted Shared VPC. This network is used to deploy an instance of the 3rd party service infrastructure for a specific health and social care customer of the 3rd party. 3rdParty-Service-Network is specific to a customer and segments their traffic from other consumers of the service. See figure 3 below for an illustration of the architecture. The model does not depict the service projects connected to each customer's restricted network.

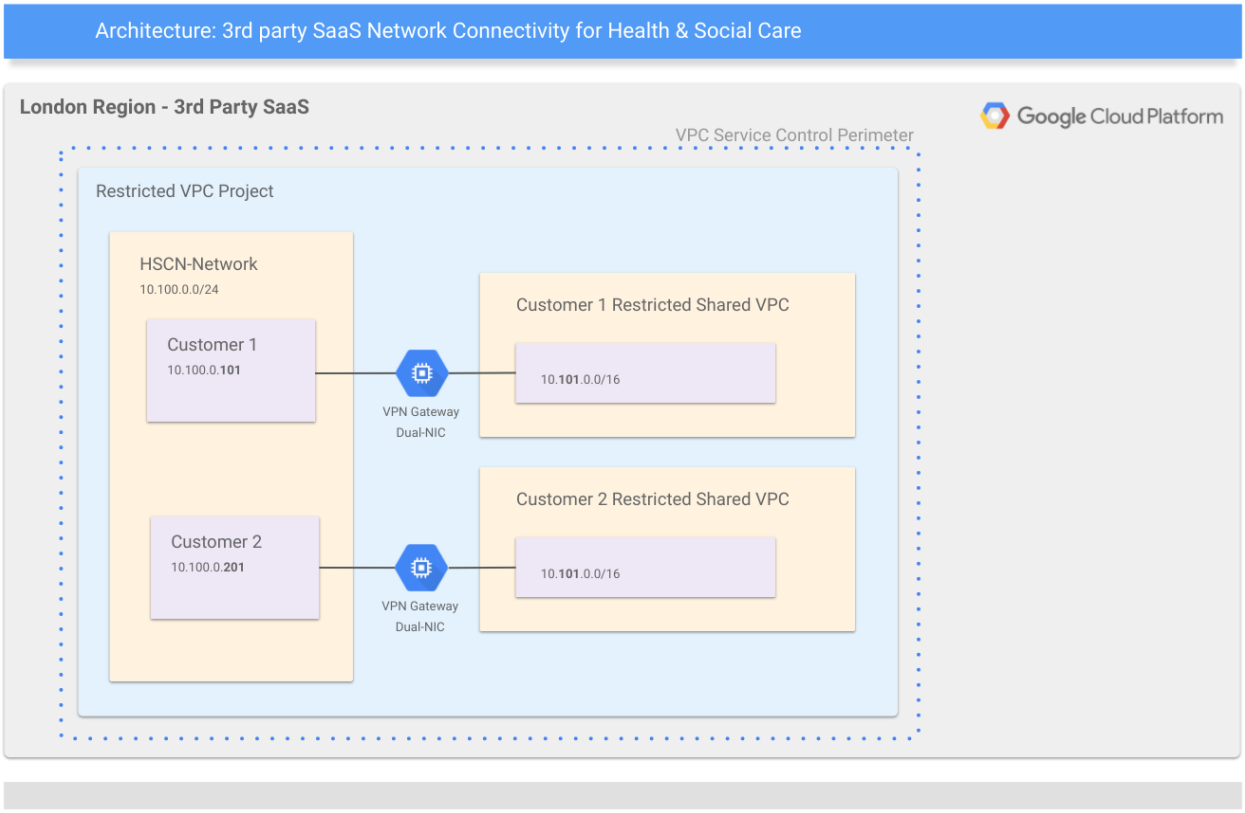


Figure 3: Architecture: 3rd party SaaS Network Connectivity for Health & Social Care

The alternative is to provision and manage dedicated links for each customer. This will increase the time required to provision any service for a customer.

Summary

Google Cloud Platform (GCP) can be used to deploy your solution in the public cloud. It can be connected to HSCN and/or your organisation as needed.

The customer is responsible for DSP Toolkit compliance and signing a HSCN Connection Agreement with a CNSP for the solution.

The network and security measures need to be carefully designed from the outset. Use the best practices provided in Google Cloud security foundations and this paper.

Use the principle of least privilege to guide your design. Setup your enterprise resource hierarchy to organise the resources and manage your policies. Separate HSCN and non HSCN facing services into separate networks and projects. Restrict access to your services using VPC service controls.

Carefully plan your IP address ranges. Use a Dual-NIC pattern to provide inbound NAT. Cloud NAT can be used to provide outbound NAT to the Internet where appropriate.

You can use the mechanisms described in this paper, including Shared VPC and VPC Peering to provide the appropriate level of isolation required for your solution.

Contact your Google Cloud account team for more detailed guidance and support in deploying HSCN solutions.

January 2021