



Hewlett Packard
Enterprise

HPE Cloud-First reference architecture guide—single DC DCN deployment

Contents

Introduction.....	2
Brownfield network.....	2
Greenfield network.....	3
DCN and Docker integration.....	4
Greenfield network design.....	4
Multiple network fabrics.....	6
External connectivity for VMs/containers.....	7
External connectivity for bare metal servers.....	8
VSD high availability.....	8
VSC high availability.....	8
VRS/5940 to VSC high availability.....	9
Out Of Band (OOB) network.....	10
SAN (Storage Area Network).....	11
DCN Deployment Overview.....	11
Product recommendations.....	12
Appendix: Sample BOM.....	14

Introduction

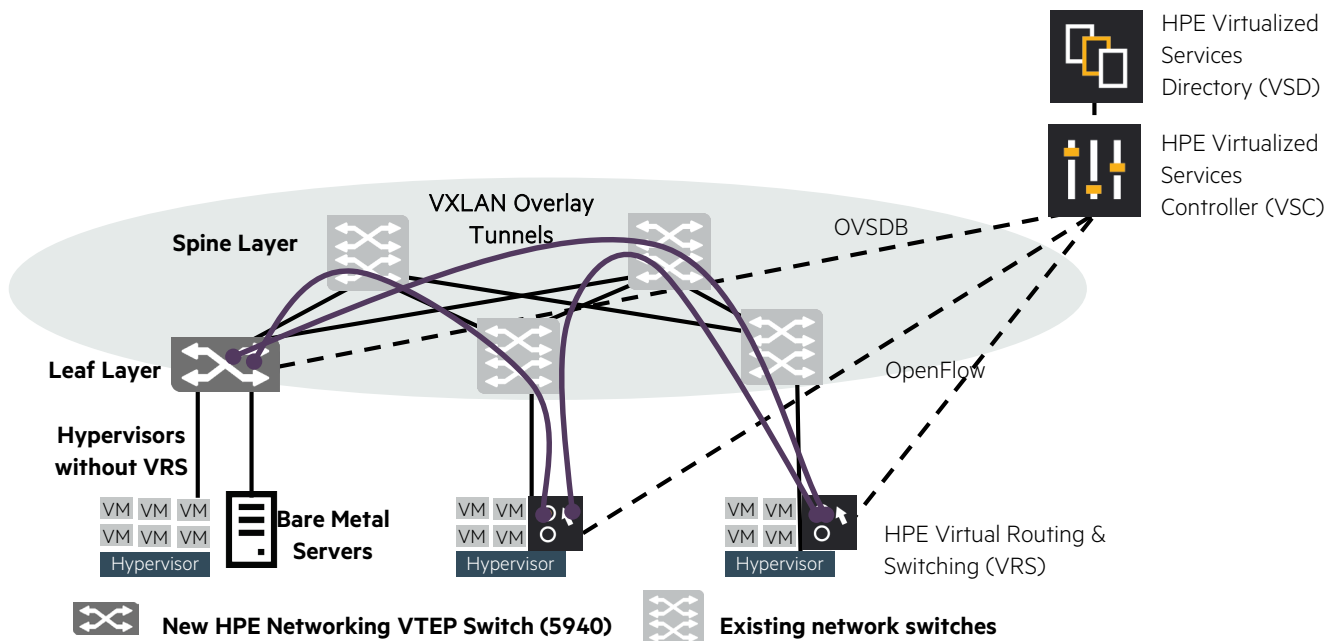
This guide provides a reference architecture for anyone interested in deploying HPE Distributed Cloud Networking (DCN) together with HPE Networking equipment in a single Data Center (DC). DCN allows service providers and large enterprises to build a simple, open and agile multi-data center environment through software-defined network (SDN) and network virtualization. HPE DCN addresses challenges such as automation, IT self-service and consistency across multiple data centers and optimization of the user experience. DCN integrates seamlessly with legacy networks, allowing you to keep your existing investments. DCN can also be deployed with new HPE Networking to provide L2/L3 network connectivity for Virtual Machines (VMs)/Docker containers with devices on the physical network. This document is based on DCN version 4 software.

Brownfield network

Customers do not need to swap out their existing network to gain the benefits of DCN. The topology as shown in figure 1 will be used to describe the brownfield network:

- The existing network (either L2 or L3) is used to provide connectivity between various DCN components such as Virtualized Services Directory (VSD), Virtualized Services Controller (VSC) and Virtual Routing & Switching (VRS)
- VRS should be installed on hypervisors that desire DCN functionality (VRS communicates with VSC via OpenFlow)
- Overlay Virtual Extensible Local Area Network (VXLAN) tunnels will be dynamically created between VXLAN Tunnel End Points (VTEPs) such as 5940 and VRS
- For hypervisors that cannot have VRS installed, they can be connected to new HPE 5940 switches functioning as VTEP to establish L2/L3 network connectivity with virtual machines connected to VRS (hypervisors without VRS will be treated as bare metal servers throughout the rest of this document)
- 5940 switches can be used when there is a requirement to provide L2/L3 network connectivity for bare metal servers to establish L2/L3 network connectivity with virtual machines connected to VRS
- 5940 switches should be deployed with Intelligent Resilient Framework (IRF) to provide chassis high availability, single IP address management, single configuration file benefits. IRF will allow servers connected to different chassis the ability to utilize active/active link aggregation for network connectivity

Figure 1. Brownfield network topology



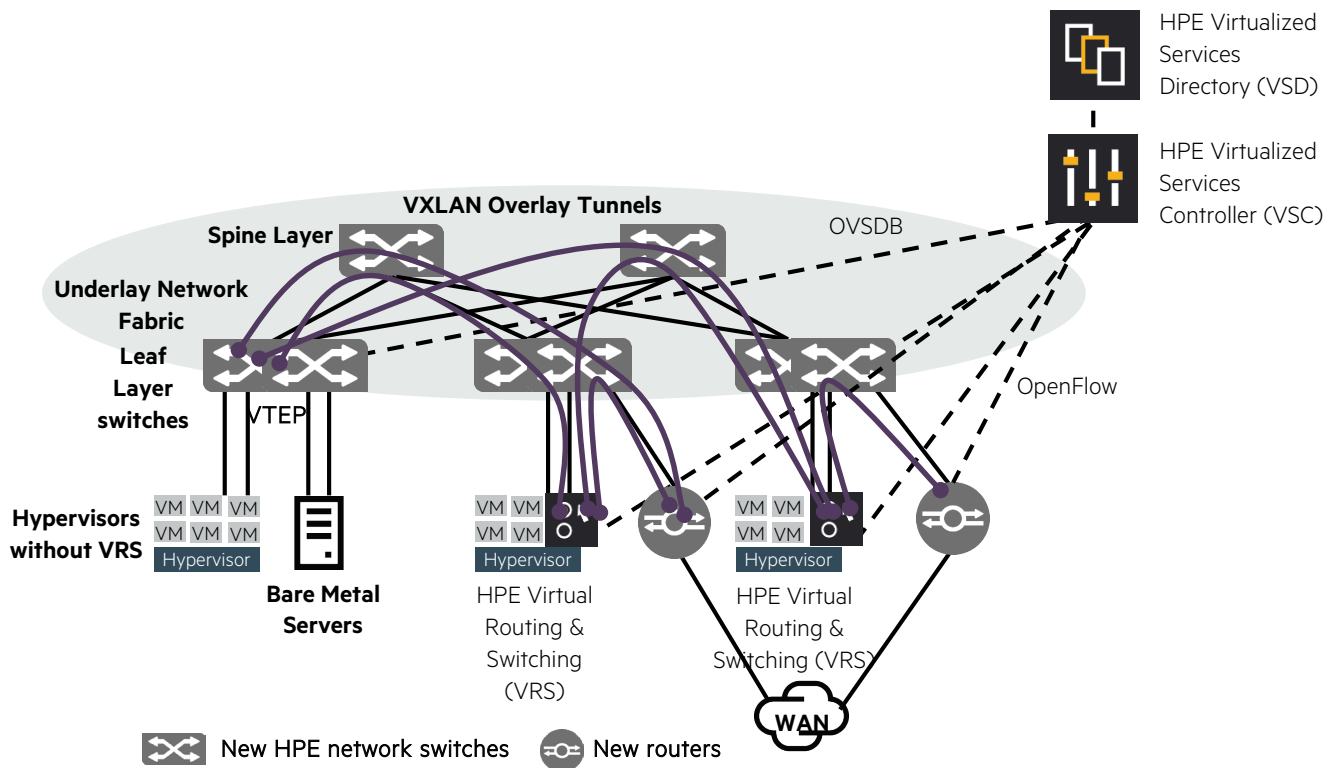
- Hypervisors with VRS do not need to be connected to VTEP enabled switches
- 5940 switches communicate with VSC via Open vSwitch Database Management Protocol (OVSDB) to share local/remote MAC addresses and dynamically create VXLAN tunnels with other VRS

Greenfield network

The topology as shown in Figure 2 will be used to describe the new greenfield network:

- HPE switches should be deployed in a CLOS leaf and spine physical topology to provide a distributed scale out, high performing, resilient Ethernet fabric with all leaf switch network ports having equal latency
- Leaf switches should be deployed with Intelligent Resilient Framework (IRF) to provide chassis high availability, single IP address management and single configuration file benefits. IRF will allow servers connected to different chassis the ability to utilize active/active link aggregation for network connectivity
- 40G links should be deployed between leaf and spine switches, these links should connect to different physical devices in an IRF switch to maximize resiliency (100G connectivity is also available on certain HPE Networking switches)
- Leaf switches can provide 10G connectivity to hypervisors or other devices e.g. WAN routers, bare metal servers
- The greenfield network will be used to transport traffic between hypervisors and various DCN components
- VRS should be installed on hypervisors that desire DCN functionality (VRS communicates with VSC via OpenFlow)
- For hypervisors that cannot have VRS installed, they can be connected to new HPE 5940 switches functioning as VTEP to establish L2/L3 network connectivity with virtual machines connected to VRS
- VTEP enabled leaf switches can be used when there is a requirement to provide L2/L3 network connectivity for bare metal servers to establish L2/L3 network connectivity with virtual machines connected to VRS

Figure 2. Greenfield network topology

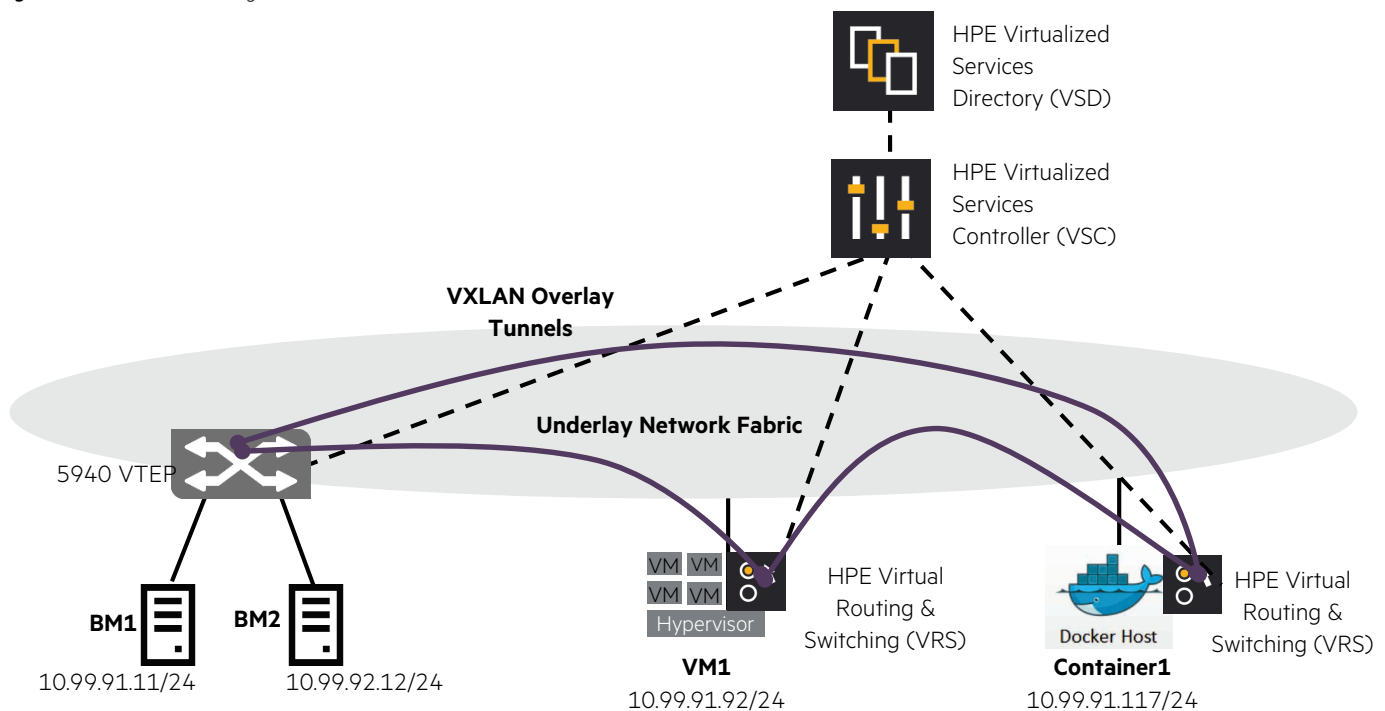


- Hypervisors with VRS do not need to be connected to VTEP enabled switches
- VTEP enabled leaf switches communicate with the VSC via Open vSwitch Database Management Protocol (OVSDB) to share local/remote MAC addresses and create VXLAN tunnels
- VTEP/OpenFlow/EVPN enabled WAN routers will connect to VRS and VTEP leaf switches for traffic to enter/exit the DC

DCN and Docker integration

DCN also supports Docker containers as shown in Figure 3. In DCN, Docker containers are treated just like VMs; this way DCN security policies are applicable to both VMs and containers at the same time. With DCN and 5940 integration, overlay network virtualization that provides L2/L3 connectivity for applications spread across bare metal servers/ VMs/ containers is now possible. To integrate Docker with DCN, VRS will need to be installed on the docker host. For the remainder of this document, any mention of hypervisors with VRS is also applicable to Docker hosts with VRS.

Figure 3. DCN and Docker integration



Greenfield network design

The topology as shown in Figure 4 is used to describe the network design:

- The underlay network design should be kept as simple as possible; this will speed up troubleshooting during outages as most of the complexity and services are moved into the servers
- An L3 IP fabric is recommended as DCN allows hypervisors to be on different subnets, this would also remove the need for Spanning Tree to be implemented
- Traceroutes and ping tests are now possible between directly connected switches to simplify underlay network troubleshooting
- Routed links between leaf and spine switches should utilize unique /30 subnets
- Rack clusters with different functionalities (e.g. Compute rack cluster/bare metal rack cluster) are recommended for ease of operations/troubleshooting, High Availability (HA) and to create separate fault domains
- One rack cluster could physically be two racks with one HPE switch in each rack or one rack with two HPE switches
- HPE switches in one rack cluster would be IRF together
- Devices in each rack cluster should utilize their directly connected IRF leaf switch as their default gateway
- With IRF switches, VRRP is not required for default gateway redundancy
- VMs and bare metal servers will utilize their local VRS as their default gateway
- Bare metal servers will utilize their local 5940s as their default gateway
- The same default gateway is programmed by VSC and distributed across 5940s and VRS as required

- Customers can start off with a two spine deployment, monitor link utilization, then add on more spine switches as required to increase bandwidth and lower oversubscription rates
- By default, Comware supports eight way ECMP = eight spine switches (modify to a higher value if necessary)
- Single area OSPF is recommended for a single network fabric (Maximum of 50 switches is a good size for one failure domain, note that 2 x physical switches forming 1 x IRF logical switch is treated as 1 x OSPF peer/router)
- DCN components such as VSD, VSC, statistics nodes and WAN routers should be placed in separate physical rack clusters for high availability (e.g. In case 1 DCN rack cluster loses power, the remaining DCN rack cluster will still be available)

Figure 4. Greenfield network topology—network design

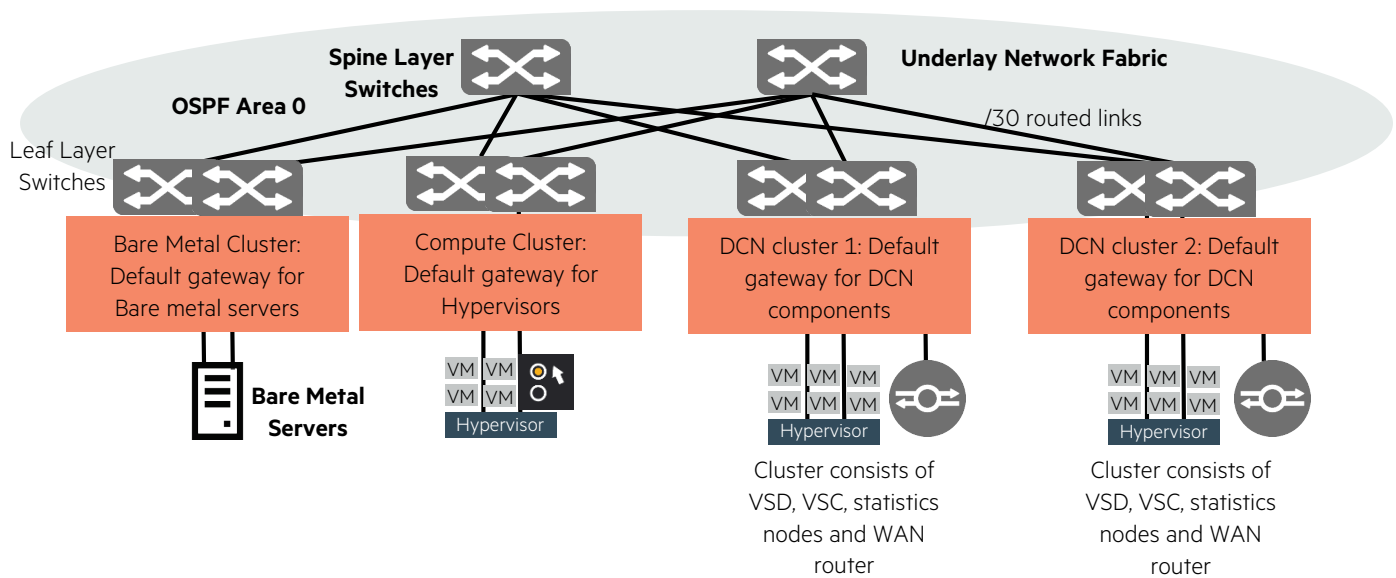
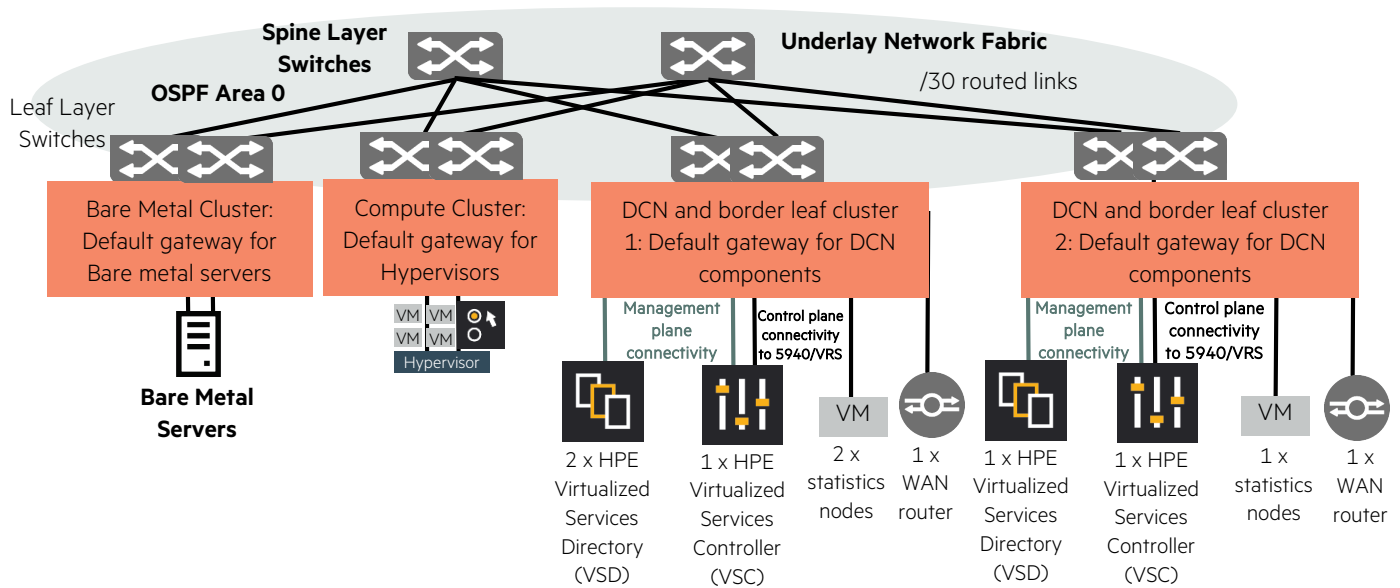


Figure 5 shows an example of the recommended deployment for DCN components across two rack clusters:

- DCN cluster 1 would have 2 x VSD, 1 x VSC, 2 x statistics nodes, 1 x WAN router
 - DCN cluster 2 would have 1 x VSD, 1 x VSC, 1 x statistics nodes, 1 x WAN router
 - Both management and control plane network connectivity would be provided to the DCN components
- VXLAN data plane would only occur between VRS/5940s

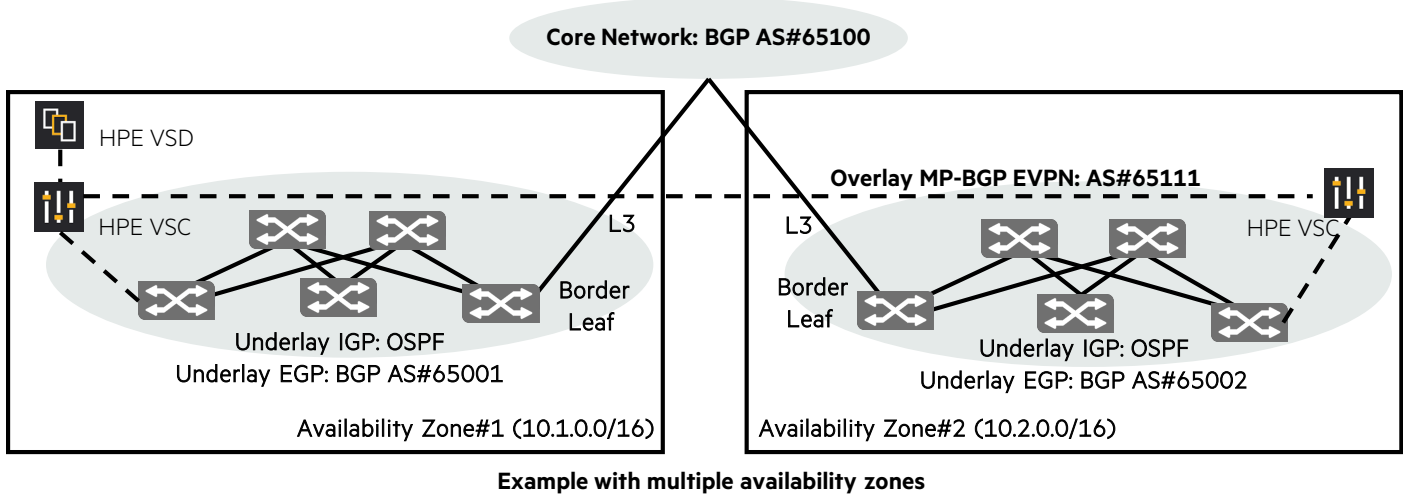
Figure 5. Greenfield network topology—DCN cluster



Multiple network fabrics

In some environments, the single DC is spread across different network fabrics/availability zones/buildings within a campus. DCN is supported across multiple fabrics/availability zones; border leaf switches should be used to connect out of an availability zone to a separate core backbone network. There should be one VSC cluster per fabric/availability zone. All VSC clusters would connect back to the main VSD cluster. These network fabrics/availability zones would typically be assigned unique private BGP AS numbers as shown in Figure 6. The overlay BGP AS and underlay BGP AS numbers would normally be different as they do not mix.

Figure 6. DCN across multiple network fabrics/availability zones

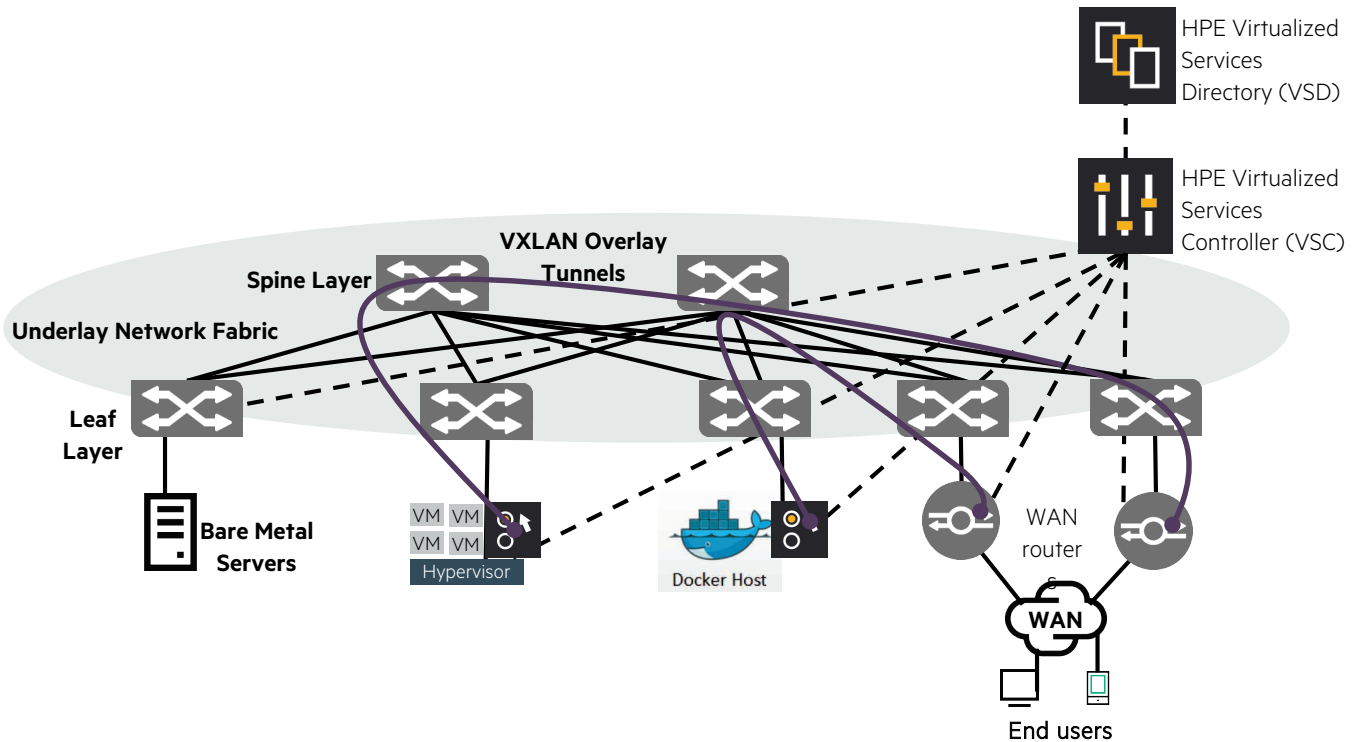


External connectivity for VMs/containers

DCN provides external connectivity from end users via the WAN using either of the following methods:

- VM/container addresses are Network Address Translated (NAT) using floating IP at VRS in both methods; this means VM/container internal IPs do not need to be exposed to the end user
- First method: DCN floating IP on compute nodes via underlay network; this means external end user routes need to be propagated into the underlay network fabric
- Second method: DCN floating IP on compute nodes via overlay to WAN routers; this means only the WAN routers need to be aware of the external end user routes as shown in Figure 7.

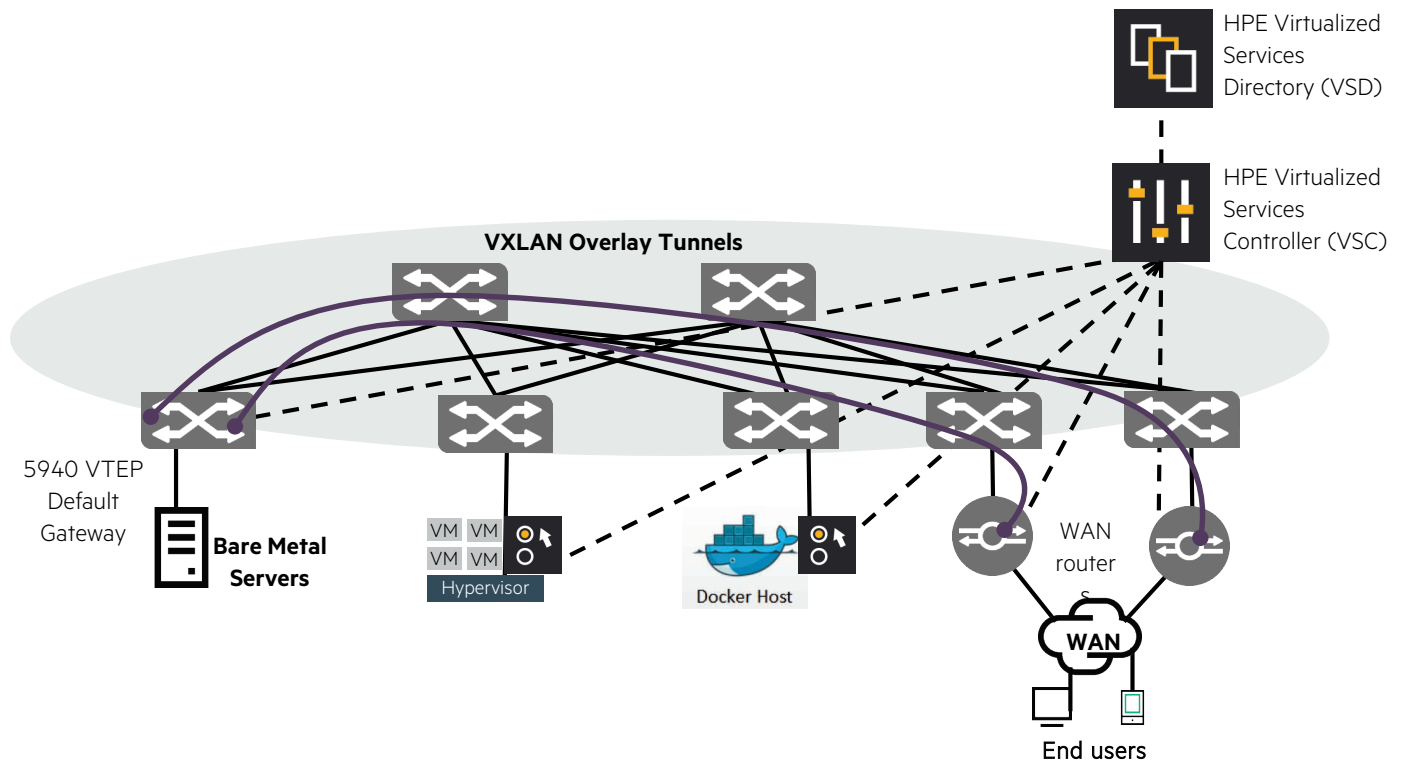
Figure 7. External connectivity for VMs/Containers



External connectivity for bare metal servers

For external end user and bare metal server connectivity, bare metal server traffic will be sent to the 5940 VTEP functioning as the default gateway to the WAN routers via VXLAN. This means only the WAN routers need to be aware of the external end user routes as shown in Figure 8.

Figure 8. External connectivity for bare metal servers



VSD high availability

As previously mentioned under Greenfield network design and shown in Figure 9, the DCN components are spread across two leaf switch rack clusters for high availability. To achieve REST API load sharing to VSDs, a traffic load balancer should be added (Refer to DCN install guide for more details).

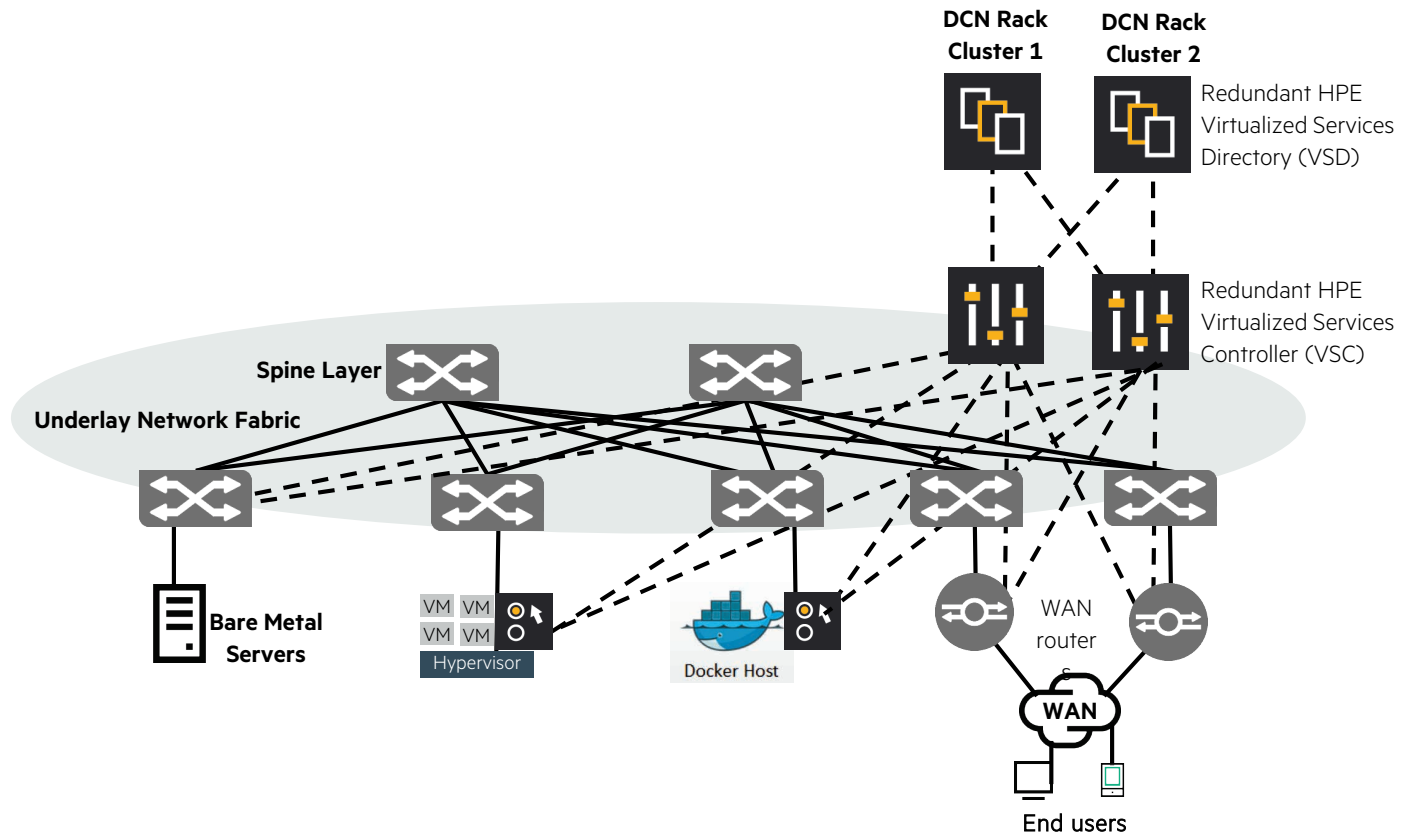
VSC high availability

Figure 9 is also applicable to VSC high availability, VSCs will support VRS within the same network fabric/availability zone. If multiple network fabrics/availability zones are deployed, separate pairs of VSCs should be placed into each availability zone/fabric. The current recommendation is for one pair of VSCs to support 500 VRS/5940s, a second pair of VSCs should be deployed in order support the next group of 500 VRS/5940s.

VRS/5940 to VSC high availability

Figure 9 is also applicable to 5940/VRS and VSC connectivity. 5940/VRS should connect to redundant VSCs within the same network fabric/availability zone. 5940/VRS will establish connectivity to both redundant VSCs. However only one VSC will function as active; the other VSC will function as backup. Load share by setting VSC1 as active for some 5940VRS, while VSC2 is active for other 5940/VRS.

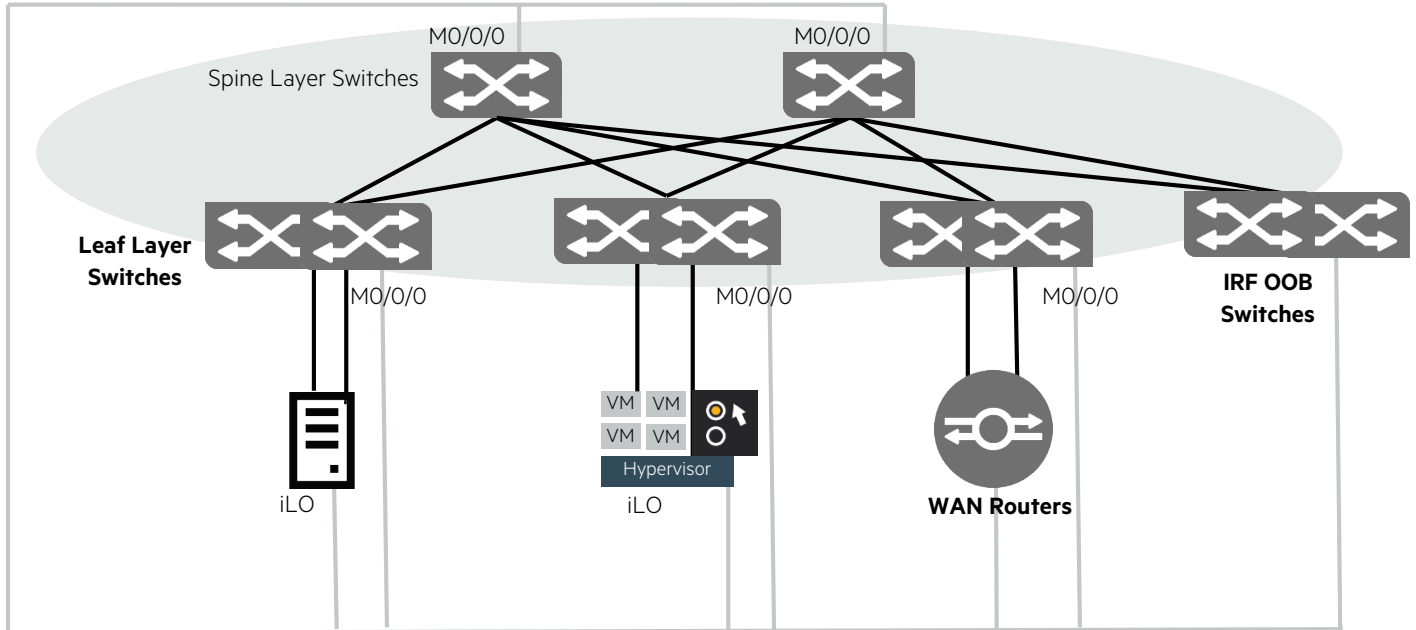
Figure 9. DCN High Availability Design



Out Of Band (OOB) network

In addition to the underlay network, an Out Of Band (OOB) network using separate OOB switches is recommended as shown in Figure 10 to provide HPE Integrated Lights-Out (iLO) connectivity if HPE servers are used. These OOB switches will also provide remote console access to switches if remote network connectivity fails via the in-band management network. The OOB switches will need to be connected to the spine switches via IGP (OSPF) for all networks to communicate.

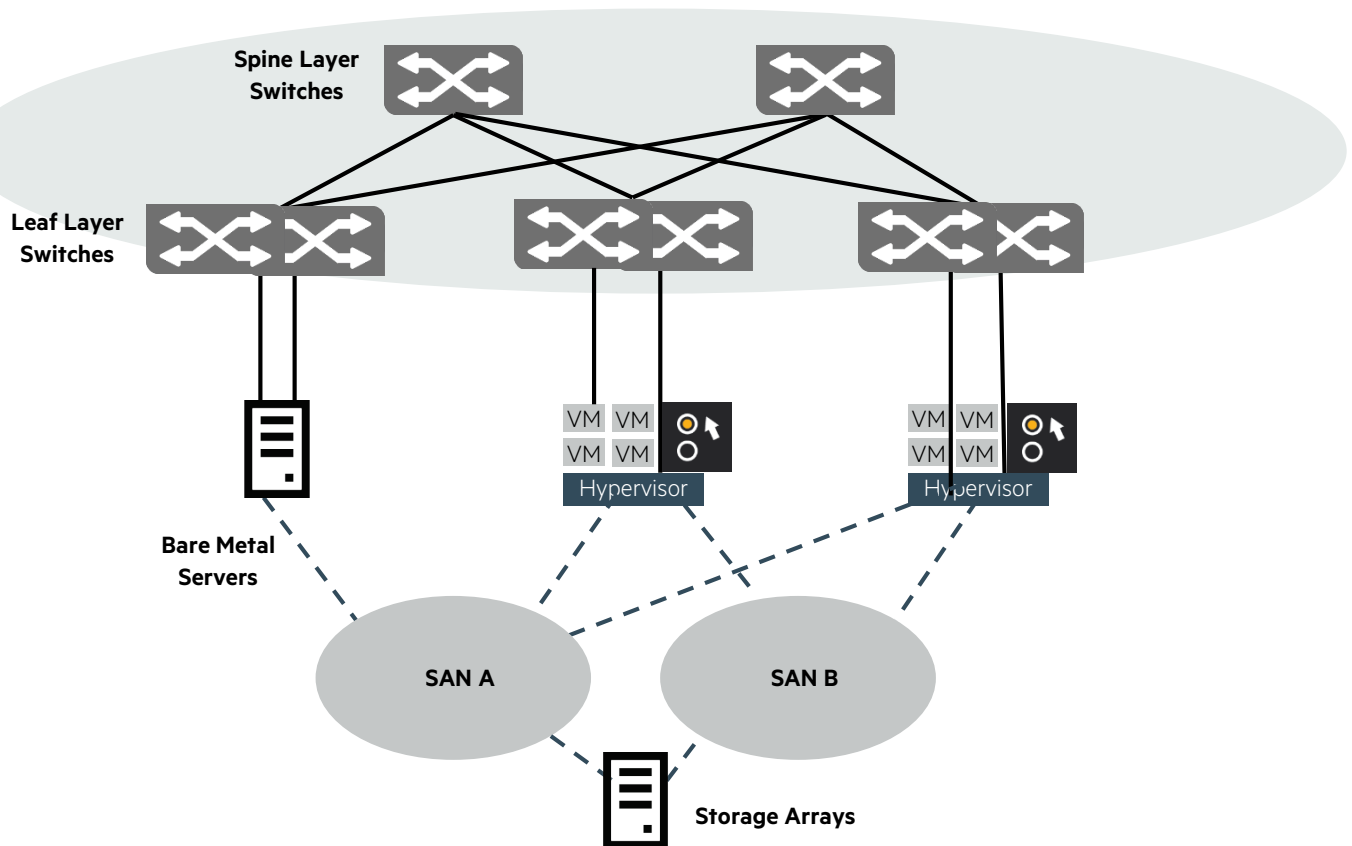
Figure 10. OOB network



SAN (Storage Area Network)

Depending on the storage array selected, a separate Fibre Channel (FC) Storage Area Network (SAN) might be required in addition to the Ethernet network fabric as shown in Figure 11 for storage traffic. Refer to SPOCK for a better understanding of the SAN deployment options available.

Figure 11. SAN



DCN Deployment Overview

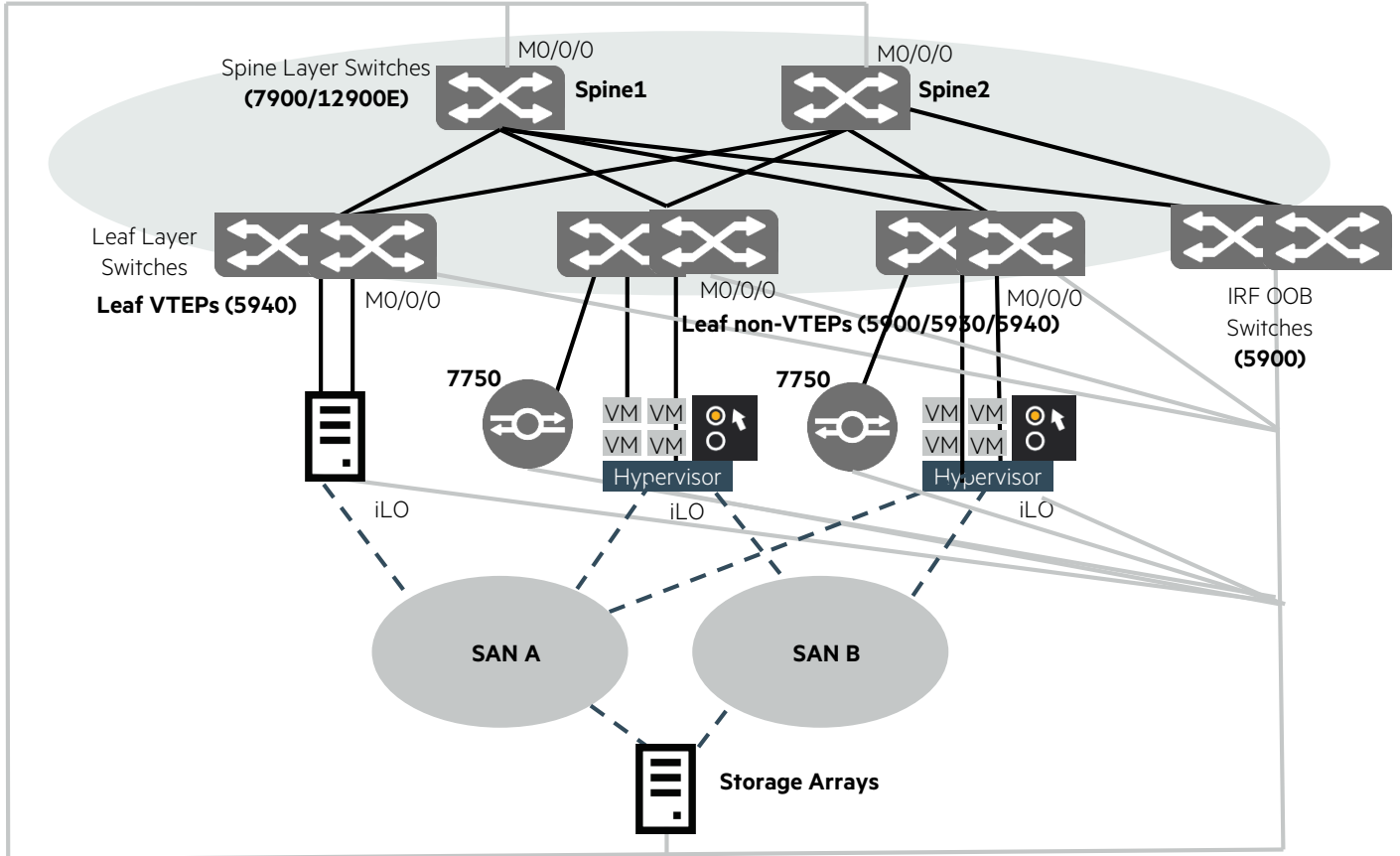
Please refer to DCN manuals (URL provided at the end of this document) for details, this section is meant to provide an overview.

- Ensure environment has NTP server, DNS server for DCN components to utilize
- If vSphere VMs need to attach to the overlay networks, DCN/vCenter integration is recommended (this integration requires an unauthenticated HTTP server and DHCP server), VMware licenses to support distributed vSwitches will also be required
- Install 3 x VSD in cluster deployment mode spread across multiple physical servers with IPs assigned to management subnet
- Install 3 x statistics node VMs with IPs assigned to management subnet
- Install 2 x VSCs with IPs assigned to both management and control/data plane subnet, configure BGP peering between VSCs and connectivity from VSC to VSDs
- Install VRS on hypervisors/Docker hosts with IPs assigned to control/data plane subnet, configure connectivity from VRS to redundant VSCs
- Docker hosts will require additional “docker monitor” software to be installed
- Assign metadata to VMs/Containers before powering them up
- Power on VMs/Containers, they should now attach to the DCN overlay networks
- Create desired network security policies between VMs/Containers

Product recommendations

The following HPE networking products as shown in Figure 12 are recommendations deemed suitable in this reference architecture.

Figure 12. Product recommendations



Spine Switch

The HPE FlexFabric 12900E Switch Series is the next generation modular data center core switch while the HPE FlexFabric 7900 Switch Series is the next generation compact modular data center core switch designed to support virtualized data centers and evolution needs of private and public cloud deployments.

Either 7900/12900E can be deployed, depending on port density/throughput/connectivity/MAC/ARP requirements to leaf switches.

Refer to product datasheets to identify port density/throughput/connectivity/MAC/ARP capabilities.

The following table provides an idea of the maximum leaf switches and total port density possible when either 7900s or 12900Es are deployed.

Table 1. Mid and large size network switch recommendations

	Greenfield network design #1 — Mid Size	Greenfield network design #2—large size
Spine switches	HPE 7900	HPE 12900E
Leaf switches	HPE 5900/5930/5940	HPE 5900/5930/5940
Maximum leaf switch count	24 to 120 IRF leaf switches	72 to 576 IRF leaf switches
Maximum port density on leaf switches	1152 to 5760 ports on leaf switches	3456 to 27648 ports on leaf switches

Leaf/OOB Switch

The HPE 5900AF series of Top of Rack (ToR) 1/10G switches brings to the market cutting edge 1/10G and 40G technologies that will advance the adoption of 10G in enterprise data centers. The HPE FlexFabric 5930 Switch Series is a family of high-density, ultra-low-latency, spine and top-of-rack (ToR) switches. The HPE FlexFabric 5940 Switch Series provides high performance and advanced features in a top-of-rack, data center switch architecture. Consisting of a 1U 32-port 100GbE QSFP28 switch, the 5940 brings high density to a small footprint. The HPE FlexFabric 5940 Switch Series is available as a 1RU delivering three configurations; 48 x 10 GbE ports with six ports of 40G, 48 x 10 GbE ports with six ports of 100G or 32 ports of 40 GbE.

5900/5930/5940 leaf switches are well suited in providing connectivity to hypervisors/servers. However, if L2/L3 VTEP functionality is required, 5940s are recommended.

SAN Switch

The HPE FlexFabric 5900CP Switch Series provides a converged, top-of-rack, data center switch architecture that offers wire once for FCoE converged environments. 5930M switches now also have the capability to add on a module to support FC/FCoE.

Refer to SPOCK for SAN details.

WAN routers

The 7750 Service Router Series delivers the performance, service richness and intelligence to drive the converged IP network edge.

Appendix: Sample BOM

This section provides some sample Bill Of Material (BOM) which should prove useful during the planning stage; Figure 2 and 12 can be referenced if required.

Spine Switch (12904E)

Table 2. Sample 12904E BOM

SKU	Description	Quantity for 2 spine switches	Comments
JH262A	HPE 12904E Switch Chassis	2	
JH108A	HPE 12900E 2400W AC PSU	8	
JH108A ABA	Included: Power Cord—U.S. localization	8	
JH265A	HPE 12904E Fan Tray Assy	4	
JH107A	HPE 12900E LPU Adapter	8	
JC665A	HPE X421 Chassis Universal 4-post RM Kit	2	
JH264A	HPE 12904E 2.5Tbps Type F Fabric Mod	12	
JH263A	HPE 12904E Main Processing Unit	4	
JH045A	HPE 12900 36p 40GbE QSFP+ FX Mod	8	
JG661A	HPE X140 40G QSFP+ LC LR4 SM XCVR	288	To leaf switches (This will be able to support 72 IRF leaves with 2 x 40G to each spine)

VTEP Leaf Switch (5940)

Table 3. Sample 5940 BOM

SKU	DESCRIPTION	QUANTITY FOR 2 LEAF SWITCHES (IRF)	COMMENTS
JH394A	HPE FF 5940 48XGT 6QSFP+ Switch	2	1 port on each switch required for IRF BFD MAD
JG552A	HPE X711 Frt(prt) Bck(pwr) HV Fan Tray	4	
JC680A	HPE 58x0AF 650W AC Power Supply	4	
JC680A ABA	included: Power Cord—U.S. localization	4	
JG661A	HPE X140 40G QSFP+ LC LR4 SM XCVR	4	2 x 40G to each spine switch

Non-VTEP Leaf Switch (5930)

Table 4. Sample 5930 BOM

SKU	Description	Quantity For 2 Leaf Switches (IRF)	Comments
JH178A	HPE 5930 2-slot 2QSFP+ Switch	2	
JC680A	HPE 58x0AF 650W AC Power Supply	4	
JC680A ABA	Included: Power Cord—U.S. localization	4	
JG552A	HPE X711 Frt(prt) Bck(pwr) HV Fan Tray	4	
JH182A	HPE 5930 24p 10GBASE-T/2p MCsc QSFP+ Mod	4	1 port on each switch required for IRF BFD MAD
JG661A	HPE X140 40G QSFP+ LC LR4 SM XCVR	4	2 x 40G to each spine switch
JG327A	HPE X240 40G QSFP+ QSFP+ 3m DAC Cable	2	Used for IRF links

Non-VTEP Leaf Switch (5900)

Table 5. Sample 5900 BOM

SKU	Description	Quantity For 2 Leaf Switches (IRF)	Comments
JG336A	HPE 5900AF 48XGT 4QSFP+ Switch	2	One port on each switch required for IRF BFD MAD
JC680A	HPE 58x0AF 650W AC Power Supply	4	
JC680A ABA	Included: Power Cord—U.S. localization	4	
JG552A	HPE X711 Frt(prt) Bck(pwr) HV Fan Tray	4	
JG661A	HPE X140 40G QSFP+ LC LR4 SM XCVR	4	2 x 40G to each spine switch
JG327A	HPE X240 40G QSFP+ QSFP+ 3m DAC Cable	2	Used for IRF links

DCN components

Table 6. Sample DCN BOM

SKU	Description	Quantity	Comments
JL025AAE	HP DCN Virtual Services Directory (VSD) Software and E-LTU	1	
JL174AAE	HP DCN VSD HA Software and E-LTU	1	
JL026AAE	HP DCN Virtual Services Controller Software and E-LTU	2	
JL177AAE	HP DCN VRS 10G KVM Software and E-LTU	1	Multiply by total number of KVM hypervisors
JL176AAE	HP DCN VRS 10G ESXi Software and E-LTU	1	Multiply by total number of ESXi hypervisors
JL027AAE	HPE VMware Plug-in for VSD SW and E-LTU	1	Only required in vSphere environments, one per vCenter

WAN routers**Table 7.** Sample DCN BOM

SKU	Description	Quantity	Comments
JL136A	7750-SR7 Starter Bundle	2	2 x 7750s in each DC
JL138A	7750 SR-7/12e AC Power Bundle	2	
JL144AAE	ALU 7750 Release 13.0 Software E-LTU	2	
JL148A	7x50 12p 10GE SFP+ Mod 8 VPRN Bndl	2	
JL156A	ALU 7x50 1p 10G SR SFP+ Transceiver	4	2 x 10G for LAN, 2 x 10G for WAN

Server**Table 8.** Sample Server BOM

SKU	Description	Quantity	Comments
818209-B21	HPE DL360 Gen9 E5-2650v4 2P 32G 8SFF Svr	1	Multiply by total number of servers required Each server includes four ports of 10Gb 533FLR-T FlexFabric Adapters
759212-B21	HP 600GB 12G SAS 15K 2.5in SC ENT HDD	1	Multiply by total number of servers required

Resources, contacts, or additional links

[HPE FlexFabric](#)

[HPE DCN Documentation](#)

[HPE DCN Manuals](#)

[HPE SPOCK](#)

Learn more at
HPE.com/networking



Sign up for updates
