

# Dell EMC Big Cloud Fabric Deployment and Best Practices Guide with VMware vSAN

Dell EMC Networking Infrastructure Solutions  
November 2017

## Revisions

| Date          | Rev. | Description   | Authors                          |
|---------------|------|---|----------------------------------|
| November 2017 | 1.01 | Corrected note in section 4.3 and minor typo corrections. | Jim Slaughter, Shree Rathinasamy |
| October 2017  | 1.0  | Initial release   | Jim Slaughter, Shree Rathinasamy |

THIS WHITE PAPER IS FOR INFORMATIONAL PURPOSES ONLY, AND MAY CONTAIN TYPOGRAPHICAL ERRORS AND TECHNICAL INACCURACIES. THE CONTENT IS PROVIDED AS IS, WITHOUT EXPRESS OR IMPLIED WARRANTIES OF ANY KIND.

Copyright © 2017 Dell Inc. All rights reserved. Dell and the Dell EMC logo are trademarks of Dell Inc. in the United States and/or other jurisdictions. All other marks and names mentioned herein may be trademarks of their respective companies.

# Table of contents

|  |    |
|--|----|
| Revisions.....   | 2  |
| 1 Introduction.....                                    | 6  |
| 1.1 Big Cloud Fabric.....                              | 7  |
| 1.2 VMware vSAN.....                                   | 8  |
| 1.3 Typographical conventions.....                     | 9  |
| 2 Hardware overview.....                               | 10 |
| 2.1 Dell EMC Networking S3048-ON.....                  | 10 |
| 2.2 Dell EMC Networking S4048-ON.....                  | 10 |
| 2.3 Dell EMC Networking Z9100-ON.....                  | 10 |
| 2.4 Dell EMC PowerEdge R740xd.....                     | 11 |
| 2.5 Dell EMC PowerEdge R630.....                       | 11 |
| 2.6 Big Switch Networks' BCF Controller Appliance..... | 11 |
| 3 Pod architectures.....                               | 12 |
| 3.1 Big Cloud Fabric pod.....                          | 12 |
| 3.2 VMware vSphere pods.....                           | 14 |
| 4 Topology.....  | 16 |
| 4.1 Production network.....                            | 16 |
| 4.2 Production network connections.....                | 17 |
| 4.3 Administrative networks.....                       | 18 |
| 4.3.1 Administrative network VLANs.....                | 19 |
| 4.3.2 OOB management network connections.....          | 20 |
| 4.3.3 P-switch control network connections.....        | 22 |
| 5 BCF deployment.....                                  | 24 |
| 5.1 BCF Controller overview.....                       | 24 |
| 5.2 Deployment overview.....                           | 25 |
| 5.3 Deployment steps.....                              | 26 |
| 5.3.1 Deploy the first BCF Controller.....             | 26 |
| 5.3.2 Deploy the second BCF Controller.....            | 29 |
| 5.3.3 Configure the cluster virtual IP address.....    | 33 |
| 5.3.4 Access the BCF GUI.....                          | 34 |
| 5.4 Switch deployment.....                             | 35 |
| 5.4.1 Zero Touch Fabric overview.....                  | 35 |

|       |   |    |
|-------|---|----|
| 5.4.2 | Collect switch MAC addresses .....                          | 36 |
| 5.4.3 | Provision switches in the BCF Controller .....              | 37 |
| 5.4.4 | Boot switches in ONIE install mode.....                     | 39 |
| 5.4.5 | Verify Switch Light OS installation.....                    | 40 |
| 5.5   | Resolve common warnings and errors .....                    | 41 |
| 5.5.1 | Suspended Switches .....                                    | 41 |
| 5.5.2 | Switches with mismatched ONIE and CPLD .....                | 41 |
| 5.5.3 | Switches without management address .....                   | 43 |
| 5.5.4 | Leaf interfaces not in interface groups .....               | 45 |
| 5.6   | BCF validation commands from the CLI.....                   | 46 |
| 5.6.1 | show fabric error .....                                     | 46 |
| 5.6.2 | show link .....   | 46 |
| 5.6.3 | show switch <i>switch name</i> interface .....              | 47 |
| 6     | VMware vSphere deployment .....                             | 48 |
| 6.1   | vCenter server deployment and design.....                   | 48 |
| 6.2   | Virtual network design .....                                | 51 |
| 6.2.1 | VDS configuration.....                                      | 52 |
| 6.2.2 | VMkernel adapter configuration.....                         | 54 |
| 7     | VMware integration with BCF .....                           | 59 |
| 7.1   | Add vCenters to BCF .....                                   | 60 |
| 7.2   | Add BCF Plugin to vCenter .....                             | 64 |
| 8     | BCF tenant and segment configuration.....                   | 66 |
| 8.1   | Overview .....  | 66 |
| 8.2   | View tenants and segments .....                             | 67 |
| 8.3   | Configure segment interfaces.....                           | 68 |
| 8.4   | Configure system tenant interfaces and logical router ..... | 71 |
| 8.5   | Verifying connectivity .....                                | 75 |
| 8.5.1 | vSAN networks .....   | 75 |
| 8.5.2 | vMotion networks.....                                       | 75 |
| 8.5.3 | VM networks .....   | 76 |
| 9     | Create vSAN clusters .....                                  | 77 |
| A     | Rack diagrams.....  | 79 |
| B     | Dell EMC validated hardware and components .....            | 80 |

|     |  |    |
|-----|--|----|
| B.1 | Switches .....   | 80 |
| B.2 | PowerEdge R740xd servers .....                                 | 80 |
| B.3 | PowerEdge R630 servers .....                                   | 81 |
| C   | Dell EMC validated software and required licenses .....        | 82 |
| C.1 | Software .....   | 82 |
| C.2 | VMware Licenses .....  | 82 |
| D   | Technical support and resources .....                          | 83 |
| D.1 | Dell EMC product manuals and technical guides .....            | 83 |
| D.2 | Big Switch Networks product manuals and technical guides ..... | 83 |
| D.3 | VMware product manuals and technical guides .....              | 83 |
| E   | Support and feedback .....                                     | 84 |

# 1 Introduction

Applications are the engines for modern businesses. They drive innovation, operational efficiency, and revenue generation. They demand an infrastructure that is highly agile and easy to manage while reducing costs. These applications, which include mission critical Enterprise Resource Planning (ERP) systems, multi-tier web applications, and big data, have placed new constraints on the networking infrastructure. Support for high east-west traffic bandwidth, virtual machine mobility, and multitenancy is critical.

Infrastructure teams have struggled to respond to these requirements. Unlike the rest of the portfolio, legacy networks remain highly static and require extensive manual intervention and operational overhead. To overcome these challenges, Software Defined Networking (SDN) is garnering due attention. SDN decouples the control plane from the data plane, allowing for dynamic management of the network. The advantages of SDN include agility, scalability, and superior network management. Open standards prevent lock-in with a single vendor and allow for financial flexibility. With such benefits, SDN solves emerging networking problems in the data center and helps keep up with virtualized environments. By providing various open networking hardware platforms and your choice of networking OS, Dell EMC Networking is the choice for future-ready data centers.

This guide covers an underlay (physical network) deployment for the Software Defined Data Center (SDDC) based on the [Dell EMC Ready Bundle for Virtualization](#) with the VMware vSAN use case. The goal of this guide is to enable a network administrator or engineer with traditional networking and VMware ESXi experience to build a scalable network using the Dell EMC Ready Bundle for Virtualization and Big Switch Networks' software outlined in this guide.

This document provides a best practice leaf-spine network topology with step-by-step configuration of a Big Cloud Fabric SDN solution integrated with VMware. It also includes an overview of the VMware distributed switches, clusters, and vSANs used in this environment.

## 1.1 Big Cloud Fabric

Dell EMC is working closely with Big Switch Networks to introduce the industry's first data center leaf-spine IP fabric solution built using Dell EMC Open Networking switches and Big Cloud Fabric (BCF). This joint solution applies the hardware-software disaggregation enabled by Dell EMC and Big Switch Networks.

With built-in integration for VMware, BCF is ideal for virtual environments, network virtualization, and Hyper-Converged Infrastructure (HCI). It is the industry's first SDN-based fabric, using Dell EMC Open Networking switch hardware that provides intelligent, agile, and flexible networking for the VMware SDDC.

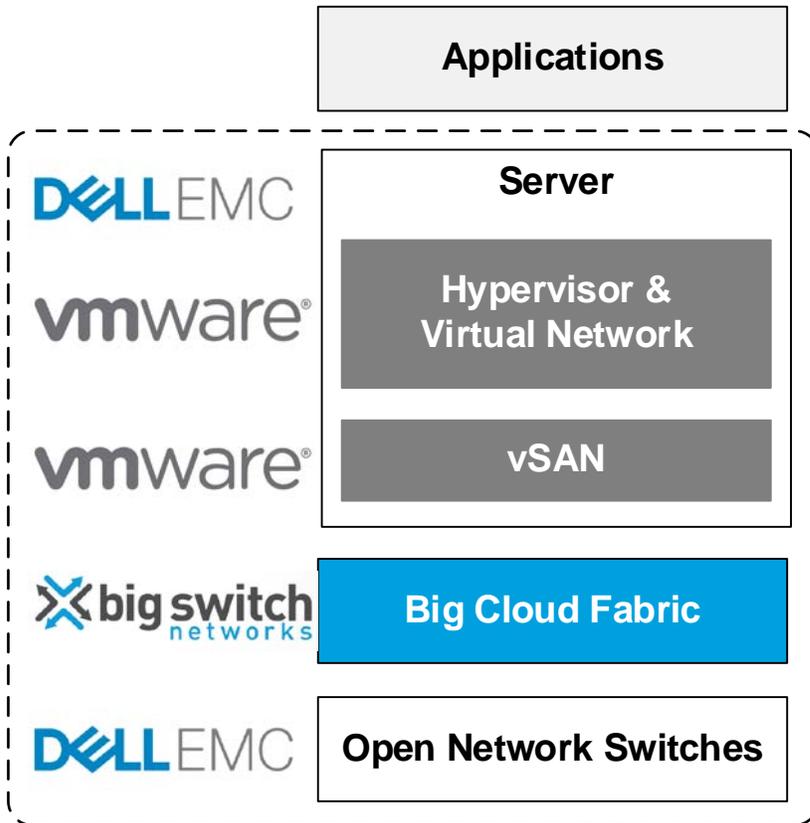


Figure 1 Dell EMC Open Networking with BCF and VMware

BCF utilizes SDN to provide scalability, improved management, visibility, flexibility, and intelligence to networks. Using redundant controllers, BCF delivers a “single logical switch” to add improved management and visibility to the network. Network agility is achieved through automation, zero-touch fabric, quicker troubleshooting, and controller-coordinated upgrades. Network automation helps in scaling seamlessly and keeping in-line with business growth and needs.

BCF allows hardware flexibility and prevents vendor lock-ins. Apart from open network hardware, BCF also helps in scaling seamlessly as per your workload needs. BCF accommodates the SDDC of the future by working in tandem with VMware vSphere, NSX, vSAN, OpenStack, VDI workloads, big data, and Software Defined Storage (SDS).

## 1.2 VMware vSAN

VMware vSAN combines the local physical storage resources of the ESXi hosts in a single cluster into a vSAN datastore. The vSAN datastore is used as the shared storage resource for creating virtual disks used by virtual machines in the cluster. vSAN is implemented directly in the ESXi hypervisor. It eliminates the need for external shared storage and simplifies storage configuration and virtual machine provisioning activities.

VMware vSphere features such as Distributed Resource Scheduling (DRS) and High Availability (HA) require shared storage. vMotion is also integrated with vSAN. vSAN provides the performance and security needed for SDDCs at a lower cost. vSAN benefits include higher performance, higher storage efficiency, scalability, ease of management, security, and automation.

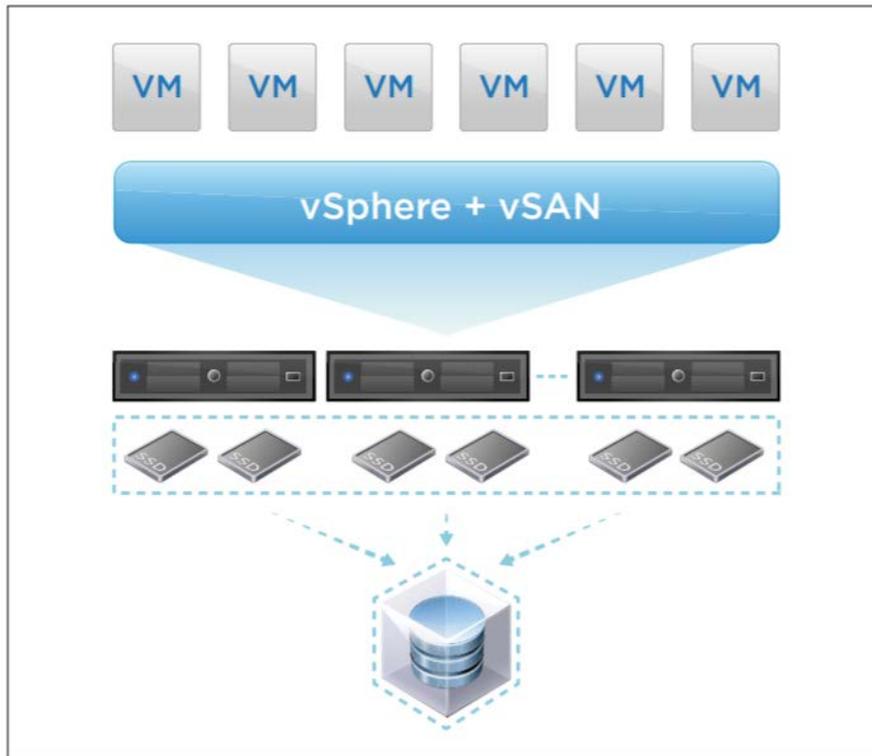


Figure 2 VMware vSAN

vSAN 6.6 features include a native HCI security solution with data-at-rest-encryption. The maintenance of vSAN is simplified with the aid of real-time support notifications and recommendations. vSAN 6.6 now uses unicast networking support, which provides easier management and setup without the need for multicast configuration. Options for automation include the vSAN SDK and PowerCLI.

For more vSAN resources, see Appendix D.

## 1.3 Typographical conventions

This document uses the following typographical conventions:

|                               |  |
|-------------------------------|--|
| Monospaced text               | Command Line Interface (CLI) examples                                    |
| <b>Bold monospaced text</b>   | Commands entered at the CLI prompt                                       |
| <i>Italic monospaced text</i> | Variables in CLI examples  |
| <b>Bold text</b>              | Graphical User Interface (GUI) fields and information entered in the GUI |

## 2 Hardware overview

This section briefly describes the hardware used to validate the deployment example in this guide. A complete listing of hardware and components used is provided in Appendix B.

### 2.1 Dell EMC Networking S3048-ON

The Dell EMC Networking S3048-ON is a 1-Rack Unit (RU) switch with forty-eight 1GbE Base-T ports and four 10GbE SFP+ ports. In this guide, one S3048-ON supports management traffic in each rack.

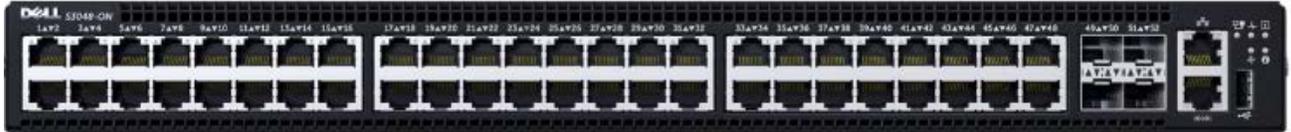


Figure 3 Dell EMC Networking S3048-ON

### 2.2 Dell EMC Networking S4048-ON

The Dell EMC Networking S4048-ON is a 1-RU, multilayer switch with forty-eight 10GbE SFP+ ports and six 40GbE QSFP+ ports. Four S4048-ON switches (two per rack) are used as leaf switches in this guide.



Figure 4 Dell EMC Networking S4048-ON

### 2.3 Dell EMC Networking Z9100-ON

The Dell EMC Networking Z9100-ON is a 1-RU, multilayer switch with thirty-two ports supporting 10/25/40/50/100GbE plus two 10GbE ports. Two Z9100-ON switches are used as spines in this guide.

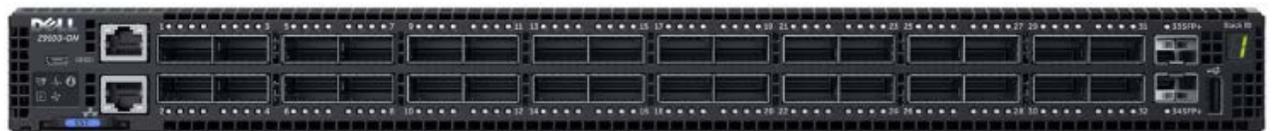


Figure 5 Dell EMC Networking Z9100-ON

## 2.4 Dell EMC PowerEdge R740xd

The Dell EMC PowerEdge R740xd is a 2-RU, two-socket server platform. It allows up to 32 x 2.5" SSDs or HDDs with SAS, SATA, and NVMe support. Ideal workloads include VMware vSANs, big data services, and data analysis. In this guide, four R740xd servers are used in the Compute cluster.



Figure 6 Dell EMC PowerEdge R740xd

**Note:** VMware recommends each vSAN cluster contain a minimum of 10% flash storage capacity. For more information, see the [VMware vSAN Design and Sizing Guide](#). The R740xd systems used in this deployment each contain twenty SSDs (100% flash storage).

## 2.5 Dell EMC PowerEdge R630

The Dell EMC PowerEdge R630 is a 1-RU, two-socket platform. In this guide, four R630 servers are used in the Management cluster.



Figure 7 Dell EMC PowerEdge R630

**Note:** For new deployments, the Dell EMC PowerEdge R640 is the latest generation 1-RU, two-socket platform and is ideal for Management cluster hosts. For existing deployments, the Dell EMC PowerEdge R630 or other supported hardware listed on the [VMware Compatibility Guide](#) web site may be used in the Management cluster.

## 2.6 Big Switch Networks' BCF Controller Appliance

The Big Switch Networks' BCF Controller Appliance is a 1-RU, two-socket platform designed to deliver the right combination of performance, redundancy, and value. For fault tolerance, two appliances are deployed into a BCF Controller cluster.



Figure 8 Big Switch Networks' BCF Controller Appliance

## 3 Pod architectures

A pod is a combination of computing, network, and storage capacity designed to be deployed as a single unit. As a result, a pod is the largest unit of failure in the SDDC. Carefully engineered services ensure each pod has little to no shared vulnerability between pods.

There are two different types of pods used in this deployment:

- Big Cloud Fabric pod
- VMware vSphere pod

### 3.1 Big Cloud Fabric pod

A BCF pod contains two BCF Controllers and a leaf-spine network spanning up to sixteen racks with two leaf switches per rack.

The sixteen-rack limit is dictated by the port count of the spine switches used in the leaf-spine network. In this deployment, Z9100-ON switches with thirty-two 40GbE interfaces per switch are used as spines. At two leaf switches per rack, all spine interfaces are used at sixteen racks.

**Note:** Big Switch Networks has tested a maximum of 128 leaf switches in one pod. See the [Big Cloud Fabric 4.2.0 Verified Scale](#) document on the Big Switch Networks' support site for more information. Big Switch Networks' documentation requires a customer account to access. Contact your Big Switch Networks account representative for assistance.

In this example, the BCF pod contains two Z9100-ON spine switches and four S4048-ON leaf switches distributed over two racks. Two BCF Controller Appliances are deployed in an active/standby configuration.

BCF provides a high bisectional bandwidth network. Each fabric device can switch at layer 2 or route at layer 3, while the BCF Controller centrally provides the intelligence required to make full use of redundant links. Incremental upgrades of the forwarding tables are dynamically pushed to each switch to ensure a stable and dynamic network operation. The Spanning Tree Protocol is not required, and all links are in forwarding mode.

The networking architecture used by BCF is a leaf-spine design that increases server-to-server bandwidth. The leaf-spine architecture creates a high-performance backplane that can be extended by simply adding more switches. Fabric edge ports are aggregated through static Multi-chassis Link Aggregation Groups (MLAGs) for higher bandwidth to all servers.

Figure 9 shows how the entire BCF pod can be thought of as a single, larger modular switch.

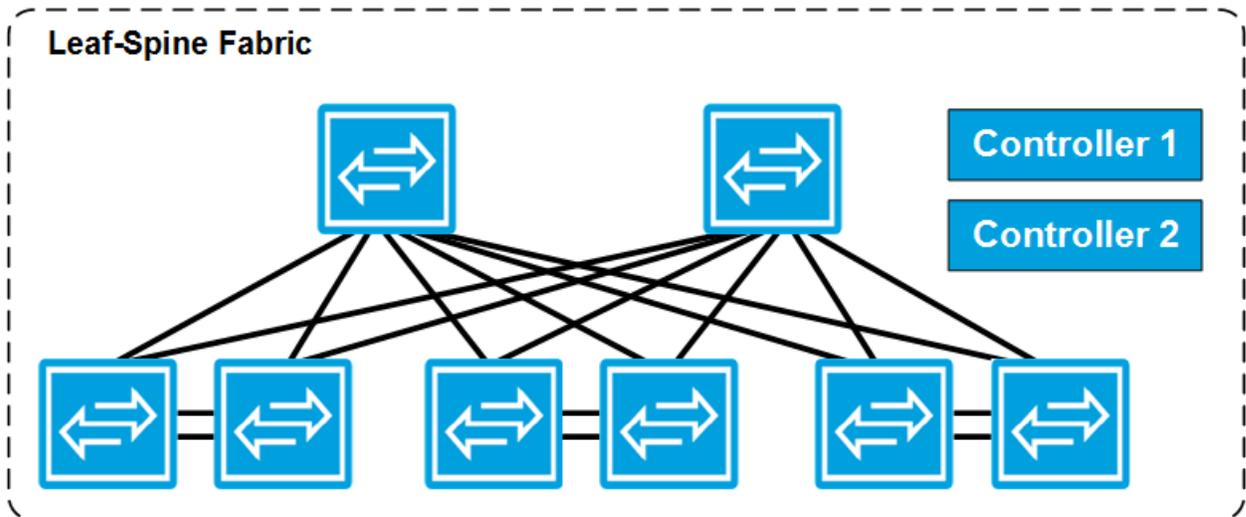


Figure 9 BCF pod

A pair of BCF Controllers provides functionality similar to the dual supervisors on a modular chassis. The spine switches provide the functionality of the backplane, while the leaf switches are similar in function to line cards.

The leaf-spine architecture provides a simple and efficient design in response to challenges inherent in the hierarchical data center architecture. The 2-layer leaf-and-spine architecture optimizes bandwidth between switch ports within the data center by creating a high-capacity fabric using multiple spine switches that interconnect the edge ports of each leaf switch. This design provides consistent latency and minimizes the hops between servers in different racks.

The design lends itself well to the creation of an independent, replicable pod that scales without disrupting network traffic. The addition of more leaf switches increases the number of switch edge ports for connecting to servers. Extra spine switches increase the fabric bandwidth and decrease oversubscription ratios.

## 3.2 VMware vSphere pods

[VMware Validated Designs Documentation](#) (VVD), release 4.1 defines the concept of VMware pods. VVD 4.1 also contains numerous best practices for VMware vSphere deployment.

There are four types of VMware vSphere pods:

- Management pod
- Compute pod
- Shared Edge and Compute pod
- Storage pod

The **Management pod** contains the hosts and virtual machines that manage the entire environment. Management, monitoring, and infrastructure services are provisioned to the Management cluster that provides high availability for these services. This includes all vCenter servers and Platform Services Controllers (PSCs) for the environment.

The **Compute pod** hosts the tenant's virtual machine (VM) workloads and vSAN storage. The pod scales by adding nodes and racks as needed, which increases computing and storage capacity linearly.

The **Shared Edge and Compute pod** uses two leaf switches to route to the core of the data center. Initially, the Shared Edge and Compute pod consists of a pair of leaf switches, shared by edge devices and computing resources. An example would be for the deployment of VMware NSX Edge Service Gateways (ESG), enabling VXLAN overlay networks.

The **Storage** pod provides secondary storage using NFS, iSCSI or Fibre Channel.

**Note:** Primary storage (vSAN storage) does not reside in the Storage pod. vSAN datastores reside with their compute resources in the Management pod, Compute pod, and Shared Edge and Compute pod.

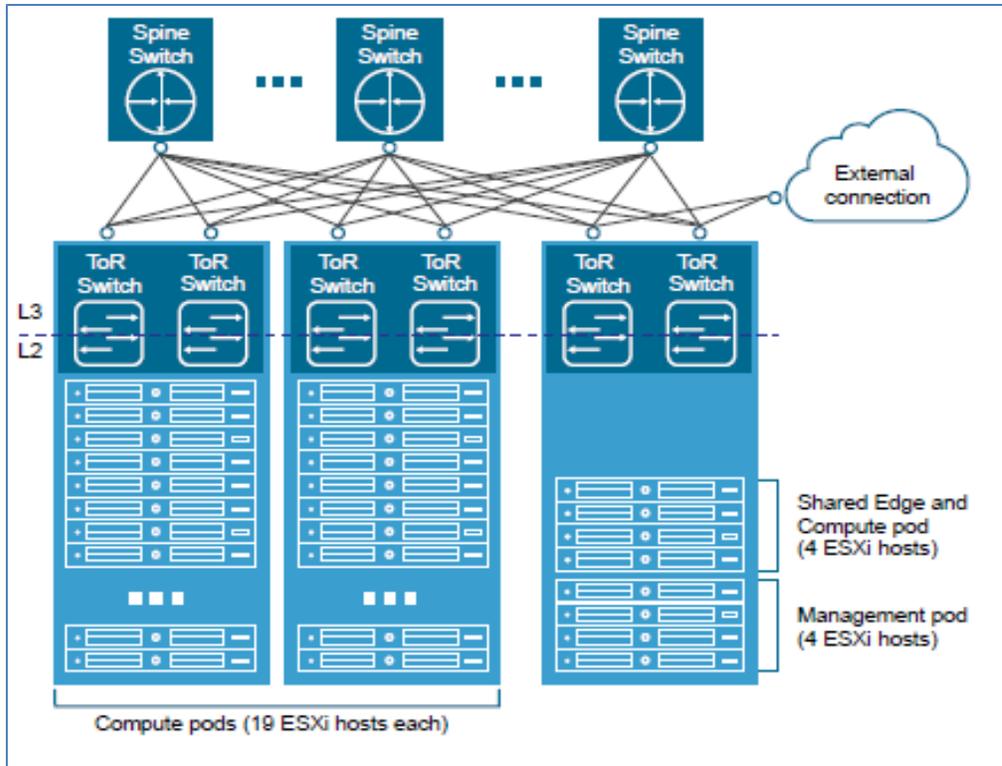


Figure 10 VMware vSphere pods

While a pod usually occupies one rack, it is possible to aggregate multiple pods into a single rack or to span a single pod across multiple racks.

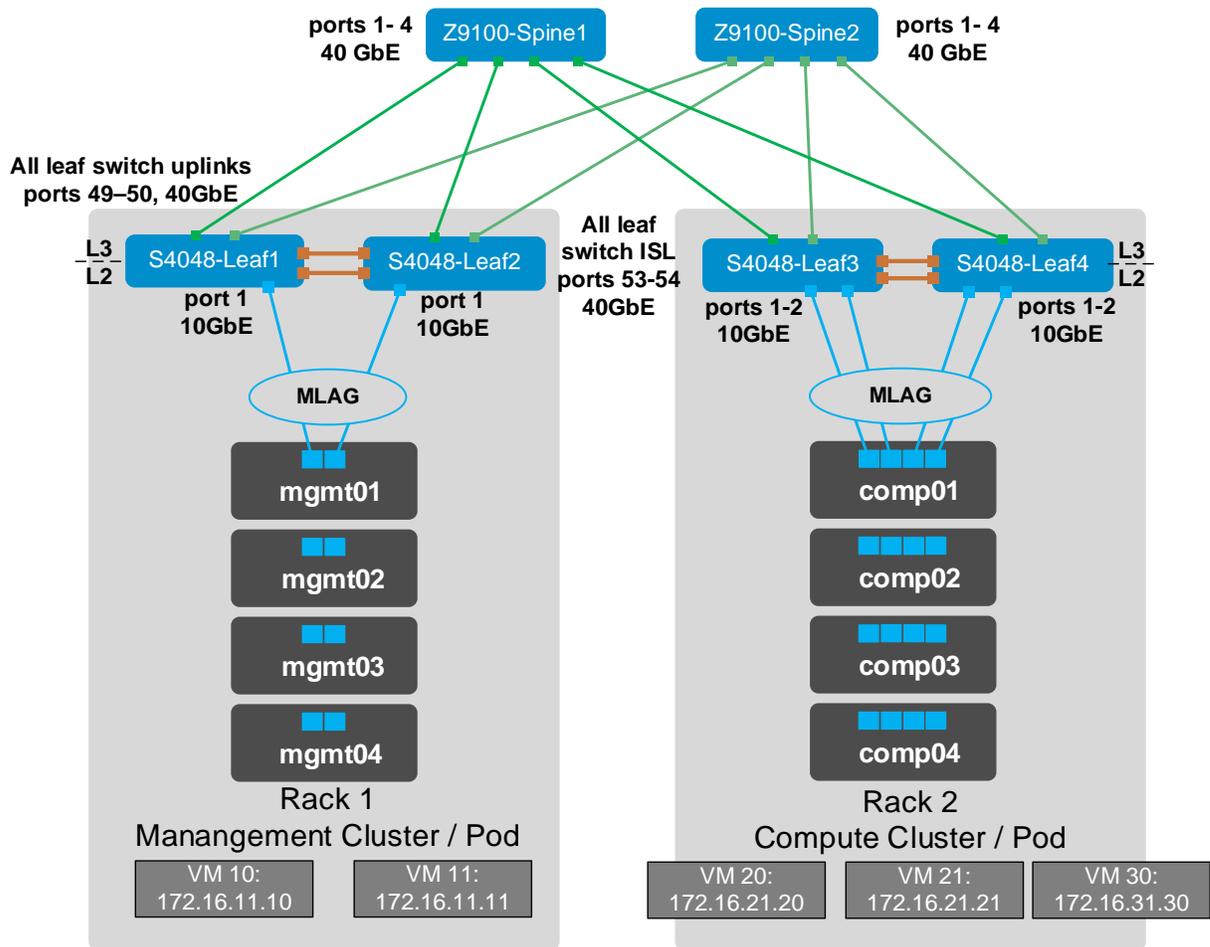
**Note:** This document covers deployment of one Management pod and one Compute pod, with each pod in one rack. Each pod is configured as a VMware cluster. Deployment of a Shared Edge and Compute pod and a Storage pod is beyond the scope of this document. See the [VMware Validated Designs Documentation](#) for deployment instructions.

## 4 Topology

The topology used in this deployment consists of a leaf-spine network for production traffic, an out-of-band (OOB) management network, and an OOB physical switch (P-switch) network for BCF Controller-to-switch communication.

### 4.1 Production network

The production network is used for VM, vSAN, and vMotion traffic. The leaf-spine topology, clusters, servers, and VMs used in this deployment are shown in Figure 11.



**Note:** Leaf connections to mgmt01 and comp01 are shown. Other hosts are connected in the same manner.

Figure 11 Production topology with leaf-spine network, ESXi hosts, and VMs

The leaf-spine topology includes two S4048-ON leaf switches at the top of each rack and two Z9100-ON spine switches. All leaf and spine switches run Big Switch Networks' Switch Light OS and are managed by BCF Controllers.

40GbE ports on each S4048-ON are used to connect to the Z9100-ON spines. Leaf switches are connected to each other with 40GbE inter-switch links (ISLs).

There are four ESXi hosts in each of the Management and Compute clusters in this example. The hosts connect to leaf switch pairs using MLAGs. The layer 2/layer 3 (L2/L3) boundary is at the leaf switches.

**Note:** Specific network interface cards (NICs) used on ESXi hosts in this paper are listed in Appendix B. See the [VMware Compatibility Guide](#) for a complete list of supported NICs.

## 4.2 Production network connections

The production network connection details for hosts and leaf switches used in this deployment are shown in Figure 12 and Figure 13. Additional hosts, not shown, are connected in the same manner.

Two 10GbE ports from each host in the Management cluster are configured in an MLAG for upstream connections to the leaf switches.

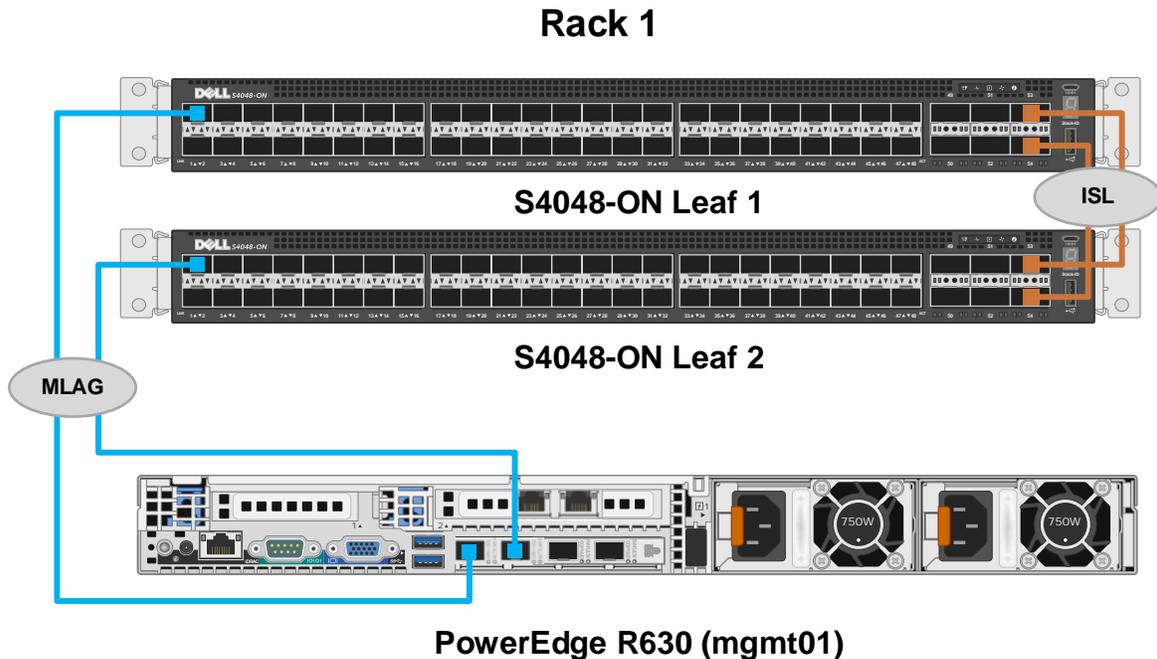


Figure 12 Production network port connection details – Management cluster hosts

In the Compute cluster, four 10GbE ports from each host are configured in an MLAG for upstream connections.

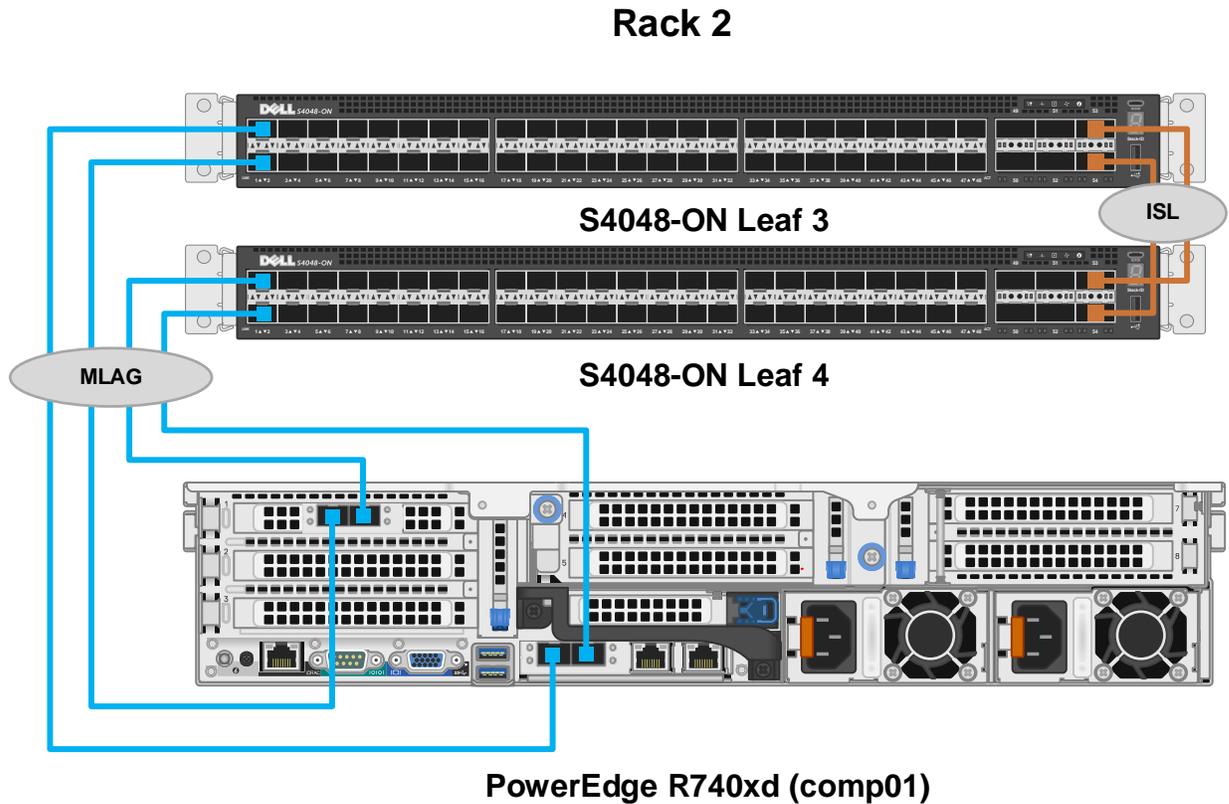


Figure 13 Production network port connection details – Compute cluster hosts

### 4.3 Administrative networks

There are two administrative OOB networks in this deployment that are isolated from the production network:

- **OOB Management network** – used for server and BCF Controller management
- **Physical switch (p-switch) control network** – used by the BCF Controller for leaf and spine switch management

One S3048-ON switch is installed as a top-of-rack (ToR) switch in each rack for administrative network traffic. Ports assigned to the two administrative networks are in separate VLANs. All connections from hosts and switches in the rack to the S3048-ON ToR switch are 1GbE Base-T.

**Note:** Using redundant S3048-ON switches in each rack is optional. Failure of either administrative network does not affect traffic on the production network. This deployment example uses a single S3048-ON switch in each rack.

### 4.3.1 Administrative network VLANs

Two VLANs are configured on each S3048-ON: VLAN 100 and VLAN 200.

Ports assigned to VLAN 100, shown in red, are used for OOB management network connections. This includes iDRAC, ESXi management, and BCF Controller management connections.

Ports assigned to VLAN 200, shown in blue, are used for p-switch control network connections.

**Note:** Big Switch Networks recommends using a dedicated broadcast domain for the p-switch control network. Using a separate VLAN accomplishes this.

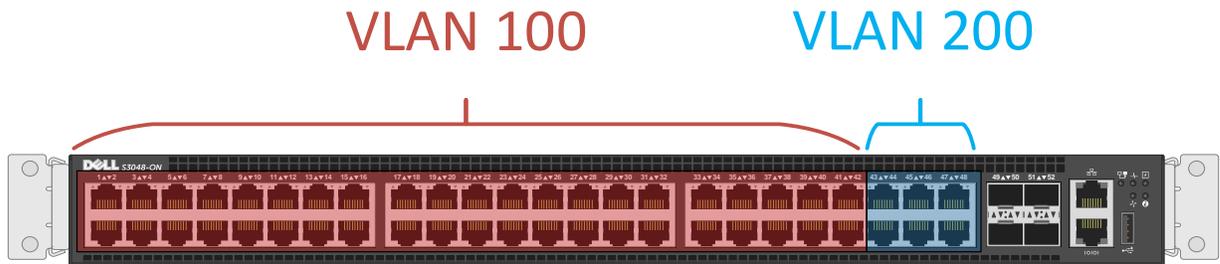


Figure 14 S3048-ON ports and VLANs

### 4.3.2 OOB management network connections

The OOB management network is used for PowerEdge server and BCF Controller configuration and monitoring. In this deployment guide, devices on the OOB management network use IPv4 addresses in the 100.67.0.0/16 address range.

As shown in Figure 15, each PowerEdge server has two connections to this network - one for ESXi management, and one for the server's iDRAC. Each BCF Controller appliance has one connection to the management network.

**Notes:** See your Dell EMC PowerEdge Server documentation for iDRAC features and instructions. BCF Controllers have two ports available for management connections. For redundancy, both controller management ports may be used, and a second S3048-ON may be added to Rack 1. See the [BCF 4.2.0 Deployment Guide](#) for more information.

All connections from hosts and switches in the rack to the S3048-ON ToR switch are 1GbE Base-T. 10GbE SFP+ ports are available on S3048-ON switches for uplinks to the management core.

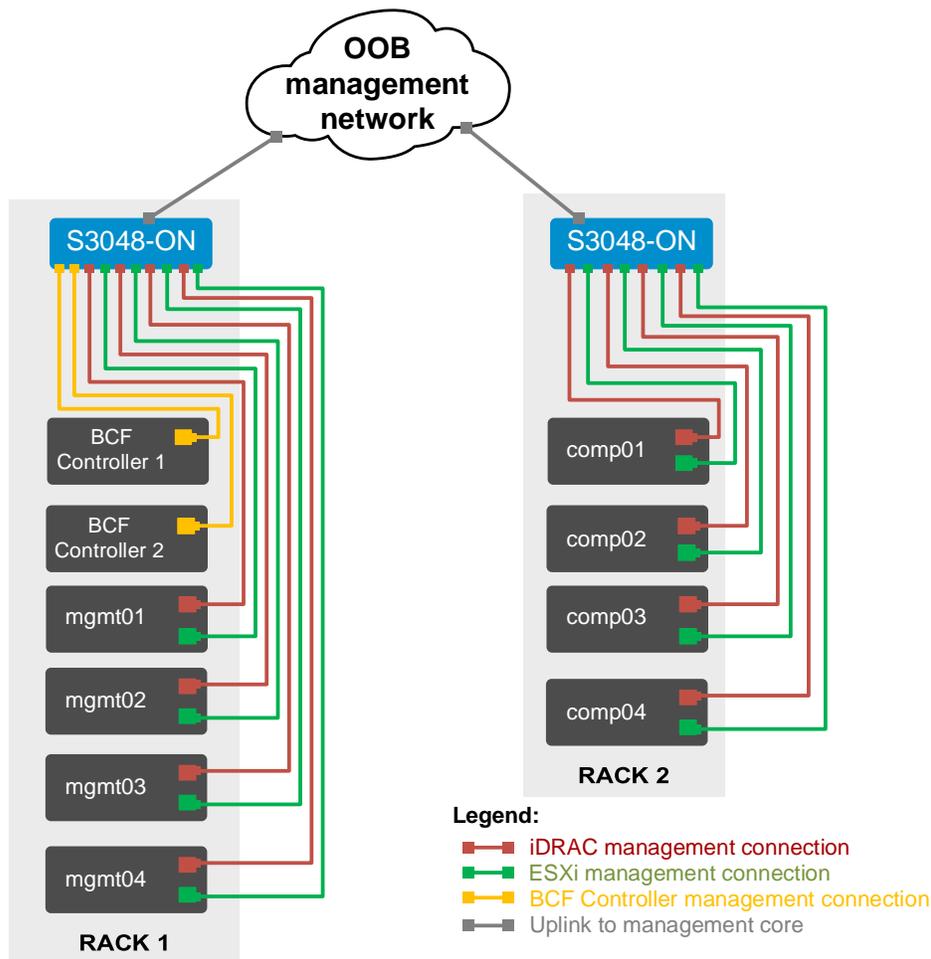


Figure 15 OOB management network connections

The OOB connection details for each unique device used in this deployment are shown in Figure 16. All connections are to ports in the OOB management VLAN (VLAN 100) on each S3048-ON switch. Additional systems, not shown, are connected in an identical manner.

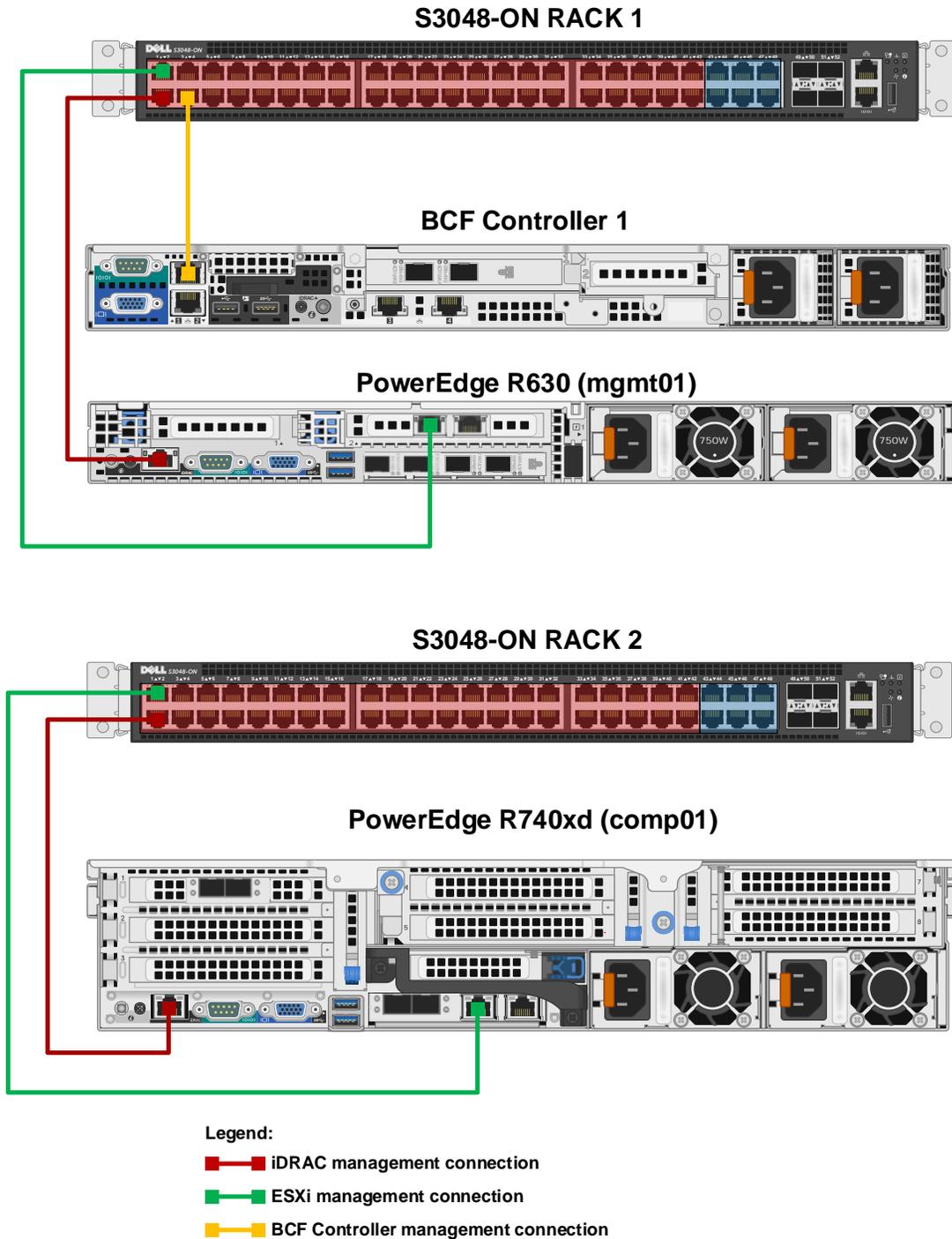


Figure 16 OOB management network port connection details

### 4.3.3 P-switch control network connections

The physical switch, or p-switch, control network contains the BCF Controllers and all leaf and spine switches. This network is used by the BCF Controllers for leaf and spine switch configuration and management.

**Note:** BCF Controllers have two ports available for p-switch control network connections. For redundancy, both ports may be used. A second S3048-ON should be added to Rack 1 for this case. See the [BCF 4.2.0 Deployment Guide](#) for more information.

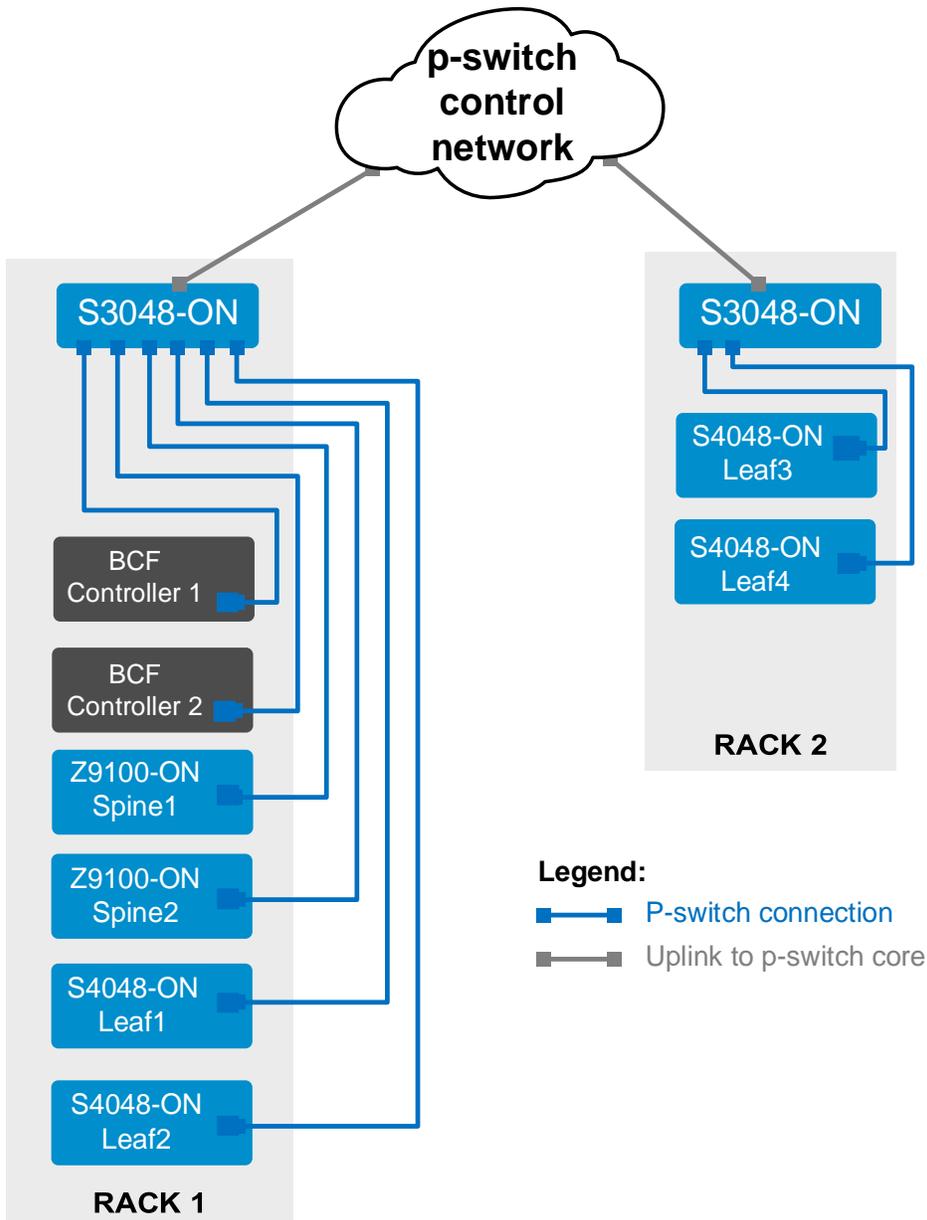


Figure 17 P-switch control network connections

In Rack 1, switch OOB management ports and BCF Controller p-switch ports are connected to S3048-ON ports in the p-switch VLAN (VLAN 200) as shown in Figure 18.

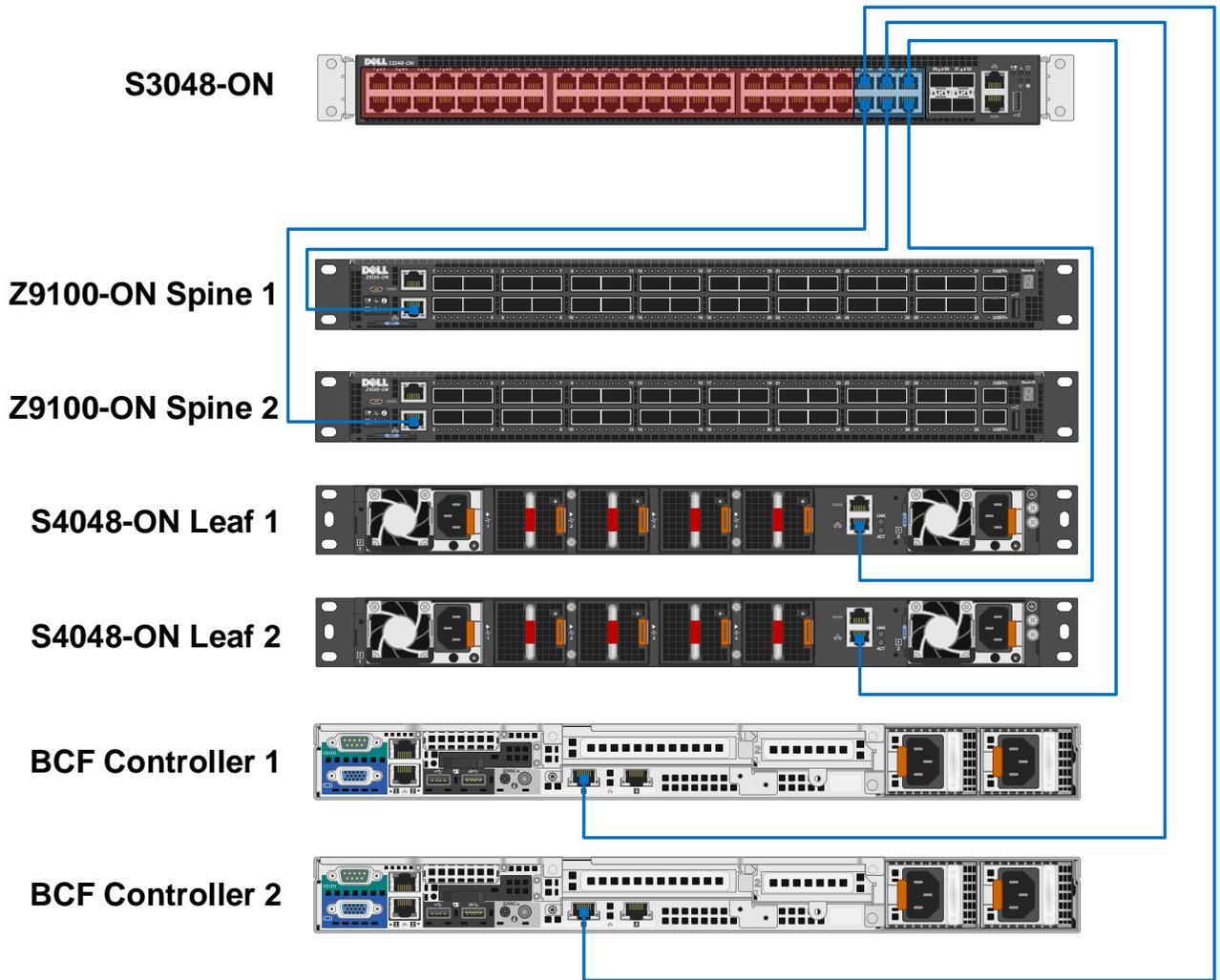


Figure 18 Rack 1 p-switch control network port connection details

The leaf switches in Rack 2, not shown, are connected in the same manner to the S3048-ON in Rack 2. The S3048-ON in Rack 2 must be able to reach the BCF Controllers via the p-switch network as shown in Figure 17.

**Note:** For small deployments or testing purposes, the leaf switches in Rack 2 may be connected directly to the S3048-ON in Rack 1. Ensure these connections are to ports in the p-switch VLAN on the S3048-ON.

# 5 BCF deployment

This section provides steps to deploy the BCF Controller cluster and how to resolve common warning and error messages.

**Note:** For more information on BCF deployment, see the [Big Cloud Fabric 4.2.0 Deployment Guide](#).

## 5.1 BCF Controller overview

The BCF Controller provides a “single pane of glass” for management of all leaf and spine switches. The BCF Controller supports a familiar CLI and a web-based GUI. Any custom orchestration can be executed by using the industry-standard RESTful application programming interface (API).

BCF supports traditional tools for debugging, including ping, traceroute, show commands, and redirecting packets using port mirroring for fault analysis. The BCF Controller also supports unique troubleshooting tools, such as Fabric Test Path and Fabric Analytics to quickly isolate, identify, and resolve forwarding and application faults.



Figure 19 BCF GUI dashboard page

## 5.2 Deployment overview

Two BCF Controllers are deployed as a cluster for redundancy. The cluster is created when the first controller is deployed as the active controller. The second controller is joined to the cluster in the standby role during deployment. If the active controller goes down, the standby controller automatically becomes the active controller.

The OOB management network settings used during deployment of each controller in this example are shown in Table 1.

**Note:** IPv4, IPv6, or both may be used on the OOB Management network. This example uses IPv4 only.

Table 1 BCF Controller initial configuration settings

| Hostname  | IP address     | IPv4 prefix length | Default gateway | DNS server address | DNS search domain |
|-----------|----------------|--------------------|-----------------|--------------------|-------------------|
| bcfctrl01 | 100.67.187.201 | 24                 | 100.67.187.254  | 100.67.189.33      | dell.local        |
| bcfctrl02 | 100.67.187.202 | 24                 | 100.67.187.254  | 100.67.189.33      | dell.local        |

Cluster settings used during deployment are shown in Table 2. A new cluster is created during deployment of the first controller. The second controller is added to the existing cluster by using the IP address of the first controller. The second controller imports the cluster name and NTP server information from the first.

Table 2 BCF Controller cluster settings

| Hostname  | Controller clustering    | Existing controller IP | Cluster name   | NTP server   |
|-----------|--------------------------|------------------------|----------------|--------------|
| bcfctrl01 | Start a new cluster      | NA                     | bcf-cluster-01 | 100.67.10.20 |
| bcfctrl02 | Join an existing cluster | 100.67.187.201         | NA             | NA           |

## 5.3 Deployment steps

This section walks through each step to set up both controllers and the cluster. The values shown in Table 1 and Table 2 are used.

### 5.3.1 Deploy the first BCF Controller

The steps to deploy the first BCF Controller are as follows:

1. Connect to the console of the first BCF Controller. The login prompt displays.

```
Big Cloud Fabric 4.2.3 (bcf-4.2.3 #35)
Log in as 'admin' to configure

controller login:
```

Figure 20 BCF login screen

2. Log in as admin (no password). **Do you accept the EULA for this product? (Yes/No/View)** displays.
3. Review the contents of the EULA if desired, and enter **Yes** to continue.
4. Enter and confirm the **Emergency recovery user password** for the controller.

```
This product is governed by an End User License Agreement (EULA).
You must accept this EULA to continue using this product.

You can view this EULA by typing 'View', or from our website at:
http://www.bigswitch.com/eula

Do you accept the EULA for this product? (Yes/No/View) [Yes] >

Running system pre-check

Finished system pre-check

Starting first-time setup

Local Node Configuration
-----

Emergency recovery user password >
Emergency recovery user password (retype to confirm) >
```

Figure 21 EULA and Emergency recovery password prompts

5. At the **Hostname>** prompt, enter the first controller's hostname, **bcfctrl01**.
6. Under **Management network options:**, select **[1] IPv4 only**.
7. Enter the values from Table 1 at the corresponding prompts:
  - a. **IPv4 address > 100.67.187.201**
  - b. **IPv4 prefix length > 24**
  - c. **IPv4 gateway > 100.67.187.254**
  - d. **DNS server 1 > 100.67.189.33**
  - e. **DNS server 2 > not used in this example**
  - f. **DNS search domain > dell.local**

The screen appears as shown in Figure 22.

```

Hostname > bcfctrl01

Management network options:

[1] IPv4 only
[2] IPv6 only
[3] IPv4 and IPv6

> 1
IPv4 address [0.0.0.0/0] > 100.67.187.201
IPv4 prefix length [24] > 24
IPv4 gateway (Optional) > 100.67.187.254
DNS server 1 (Optional) > 100.67.189.33
DNS server 2 (Optional) >
DNS search domain (Optional) > dell.local

```

Figure 22 Hostname and OOB management network settings

8. Under **Controller cluster options:**, select **[1] Start a new cluster**.
9. Enter the **Cluster name**, **bcf-cluster-01**.
10. Enter a **Cluster description (Optional)**. A description is not used in this example.
11. Enter a **Cluster administrator password** and retype to confirm.

The screen appears as shown in Figure 23.

```

Controller Clustering
-----

Controller cluster options:

[1] Start a new cluster
[2] Join an existing cluster

> 1
Cluster name > bcf-cluster-01
Cluster description (Optional) >
Cluster administrator password >
Cluster administrator password (retype to confirm) >

```

Figure 23 Create a new cluster

12. Under **NTP server options:**, select option [1] (default) or [2]. In this example, [2] **Use Custom NTP servers** is selected, and the **NTP server 1** address used is **100.67.10.20**. **NTP server 2 (Optional)** is not used in this example.

```
System Time
-----

Default NTP servers:

- 0.bigswitch.pool.ntp.org
- 1.bigswitch.pool.ntp.org
- 2.bigswitch.pool.ntp.org
- 3.bigswitch.pool.ntp.org

NTP server options:

[1] Use default NTP servers
[2] Use custom NTP servers

[1] > 2
NTP server 1 > 100.67.10.20
```

Figure 24 NTP server selection

13. A summary of the configuration settings displays. Review the settings and select [1] **Apply settings**.

```
Menu
----

Please choose an option:

[ 1] Apply settings
[ 2] Reset and start over
[ 3] Update Recovery Password      (*****)
[ 4] Update Hostname              (bcfctr101)
[ 5] Update IP Option             (IPv4 only)
[ 6] Update IPv4 Address          (100.67.187.201)
[ 7] Update IPv4 Prefix Length    (24)
[ 8] Update IPv4 Gateway          (100.67.187.254)
[ 9] Update DNS Server 1         (100.67.189.33)
[10] Update DNS Server 2         (<none>)
[11] Update DNS Search Domain    (dell.local)
[12] Update Cluster Option       (Start a new cluster)
[13] Update Cluster Name         (bcf-cluster-01)
[14] Update Cluster Description  (<none>)
[15] Update Admin Password       (*****)
[16] Update NTP Option           (Use custom NTP servers)
[17] Update NTP Server 1         (100.67.10.20)
[18] Update NTP Server 2         (<none>)

[1] >
```

Figure 25 Configuration summary – first BCF Controller

14. The settings are applied and the **First-time setup is complete!** message displays.

```
[1] > 1
[Stage 1] Initializing system
[Stage 2] Configuring controller
  Waiting for network configuration
  IP address on bond0 is 100.67.187.201
  Generating cryptographic keys
[Stage 3] Configuring system time
  Initializing the system time by polling the NTP server:
  100.67.10.20
[Stage 4] Configuring cluster
  Cluster configured successfully.
  Current node ID is 13684
  All cluster nodes:
  Node 13684: fe80::1618:77ff:fe5b:3cc3:6642

First-time setup is complete!

Press enter to continue >
```

Figure 26 Configuration settings applied

15. Press **Enter**. The controller hostname and login prompt displays.

```
Big Cloud Fabric 4.2.3 (bcf-4.2.3 #35)
Log in as 'admin' to configure

Hint: Num Lock on

bcfctr101 login:
```

Figure 27 Controller login prompt

### 5.3.2 Deploy the second BCF Controller

Setting up the second controller is similar to the first, except that the second controller joins the existing cluster configured on the first controller.

1. Connect to the console of the second BCF Controller. The login prompt displays as shown in Figure 20 in the previous section.
2. Log in as **admin** (no password), accept the EULA, and provide/confirm the **Emergency recovery user password**.
3. At the **Hostname>** prompt, enter the second controller's hostname, **bcfctr102**.
4. Under **Management network options:**, select **[1] IPv4** only.
5. Enter the values from Table 1 at the corresponding prompts:
  - a. **IPv4 address > 100.67.187.202**
  - b. **IPv4 prefix length > 24**
  - c. **IPv4 gateway > 100.67.187.254**
  - d. **DNS server 1 > 100.67.189.33**
  - e. **DNS server 2 > not used in this example**
  - f. **DNS search domain > dell.local**

The screen appears as shown in Figure 28.

```
Hostname > bcfctr102

Management network options:

[1] IPv4 only
[2] IPv6 only
[3] IPv4 and IPv6

> 1
IPv4 address [0.0.0.0/0] > 100.67.187.202
IPv4 prefix length [24] > 24
IPv4 gateway (Optional) > 100.67.187.254
DNS server 1 (Optional) > 100.67.189.33
DNS server 2 (Optional) >
DNS search domain (Optional) > dell.local
```

Figure 28 Management network configuration – second BCF Controller

6. Under **Controller cluster options:**, select **[2] Join an existing cluster**.
7. Enter the **Existing controller address, 100.67.187.201**. This is the IP address of the first controller.
8. Enter the **Cluster administrator password** that was configured on the first controller and retype to confirm.

The screen appears as shown in Figure 29.

```
Controller Clustering
-----

Controller cluster options:

[1] Start a new cluster
[2] Join an existing cluster

> 2
Existing controller address > 100.67.187.201
Cluster administrator password >
Cluster administrator password (retype to confirm) >
```

Figure 29 Joining an existing cluster

9. The configuration settings summary for the second controller displays. Review the settings and select **[1] Apply settings**.

```
Menu
----
Please choose an option:

[ 1] Apply settings
[ 2] Reset and start over
[ 3] Update Recovery Password      (*****)
[ 4] Update Hostname               (bcfctr102)
[ 5] Update IP Option              (IPv4 only)
[ 6] Update IPv4 Address           (100.67.187.202)
[ 7] Update IPv4 Prefix Length     (24)
[ 8] Update IPv4 Gateway           (100.67.187.254)
[ 9] Update DNS Server 1          (100.67.189.33)
[10] Update DNS Server 2          (<none>)
[11] Update DNS Search Domain     (dell.local)
[12] Update Cluster Option         (Join an existing cluster)
[13] Update Existing Controller   (100.67.187.201)
[14] Update Admin Password        (*****)

[1] >
```

Figure 30 Configuration summary on second BCF Controller

10. Once the settings are applied, the screen appears as shown in Figure 31. The message **Please verify that: Secure control plane is NOT configured** displays. By default, the secure control plane is not configured.

```
[1] > 1
[Stage 1] Initializing system
[Stage 2] Configuring controller
  Waiting for network configuration
  IP address on bond0 is 100.67.187.202
  Generating cryptographic keys

Please verify that:

  Secure control plane is NOT configured.

You can verify the above by running "show secure control plane"
on the existing controller 100.67.187.201.

Options:

[1] Continue connecting (the above info is correct)
[2] Cancel and review parameters

>
```

Figure 31 Applying settings on second controller

**Note:** The secure control plane is a feature where certificates are issued by a trusted Certificate Authority (CA) to each controller and switch that will participate in the fabric. When enabled, controllers and switches cannot join the fabric without a valid certificate from the trusted CA. The secure control plane is not configured by default. See the [BCF 4.2.0 User Guide](#) for more information on this feature.

11. Select option **[1] Continue connecting (the above info is correct)** to proceed. When done, the message **First-time setup is complete!** is displayed.

```
Options:
[1] Continue connecting (the above info is correct)
[2] Cancel and review parameters
> 1

[Stage 3] Configuring system time
  Initializing the system time by polling the NTP server:
    100.67.10.20
[Stage 4] Configuring cluster
  Cluster configured successfully.
  Current node ID is 20702
  All cluster nodes:
    Node 13684: fe80::1618:77ff:fe5b:3cc3:6642
    Node 20702: fe80::b283:feff:fed6:3e8f:6642

First-time setup is complete!

Press enter to continue >
```

Figure 32 Settings applied on second controller

12. Press the **Enter** key. The controller login screen for the second controller displays.
13. Log in as **admin**. The command prompt displays and indicates this controller is in the **standby** role.

```
Big Cloud Fabric 4.2.3 (bcf-4.2.3 #35)
Log in as 'admin' to configure

Hint: Num Lock on

bcfctr102 login: admin
Password:
Last login: Mon Oct 16 18:48:19 UTC 2017 on tty1
Big Cloud Fabric 4.2.3 (bcf-4.2.3 #35)
Logged in as admin, 2017-10-16 20:14:03.192000 UTC, auth from bcfctr102
standby bcfctr102>
```

Figure 33 Login prompt and command prompt on second (standby) controller

**Note:** The standby controller is read only. Configuration commands made from the command line must be run on the active controller.

### 5.3.3 Configure the cluster virtual IP address

As a best practice, set a virtual IP (VIP) address for the cluster. This allows you to connect to the management port of the active node using an IP address that does not change even if the active controller fails over and the role of the standby controller changes to the active.

1. Log in to the console of the active controller locally or remotely using secure shell (SSH).
2. Use the set of commands shown in Figure 34 to set the cluster VIP address.

```
bcfctr101> enable
bcfctr101# configure
bcfctr101(config)# controller
bcfctr101(config-controller)# virtual-ip 100.67.187.200
```

Figure 34 Setting the cluster virtual IP address

3. To verify the cluster settings, enter the `show controller` command from either the active or standby controller. Verify that the **Cluster Virtual IP** address is correct and that **Redundancy status** is **redundant**.

```
bcfctr101(config-controller)# show controller
Cluster Name           : bcf-cluster-01
Cluster Virtual IP     : 100.67.187.200
Redundancy Status     : redundant
Last Role Change Time : 2017-10-16 18:38:37.414000 UTC
Failover Reason       : Changed connection state: cluster configuration changed
Cluster Uptime        : 2 hours, 22 minutes
# IP                  @ State  Uptime
-|-----|-----|-----|
1 100.67.187.201 * active 2 hours, 22 minutes
2 100.67.187.202  standby 50 minutes
```

Figure 35 Show controller command output

### 5.3.4 Access the BCF GUI

The BCF GUI is accessible from a browser using the VIP address of the cluster.

**Note:** For more information, see the [Big Cloud Fabric 4.2.0 GUI Guide](#).

1. Enter the cluster VIP address in a web browser. You are redirected to a secure login page.

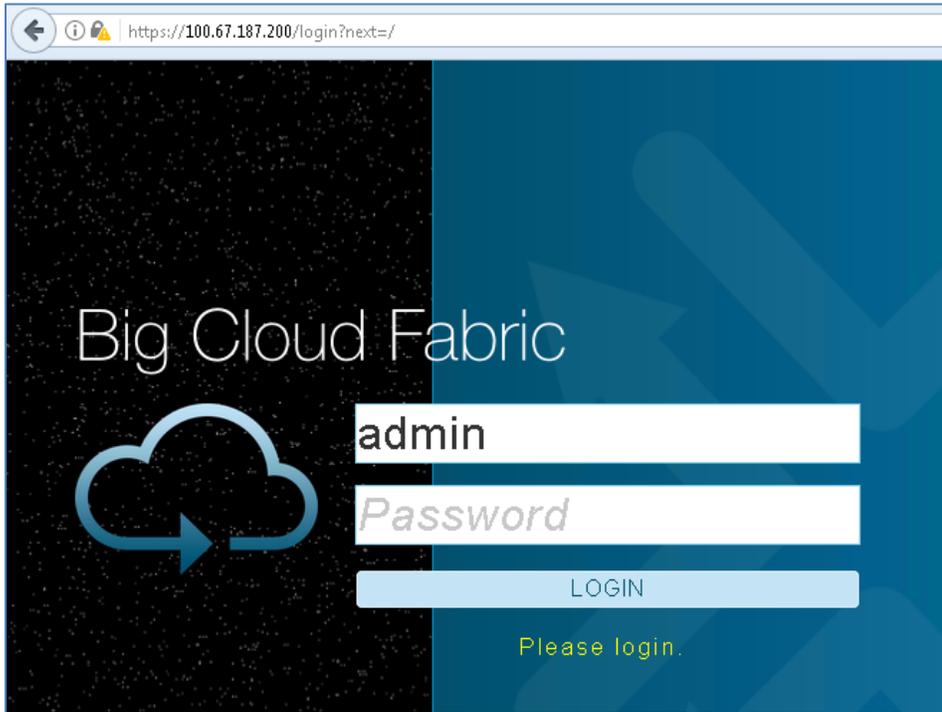


Figure 36 Connecting to the BCF GUI

**Note:** The BCF controller uses a self-signed certificate. See the [Big Cloud Fabric 4.2.0 User Guide](#) to install a certificate from a trusted CA.

2. Log in as **admin** using the password created during controller setup. The BCF dashboard displays with **Attributes** and **Controller Stats** similar to Figure 37.

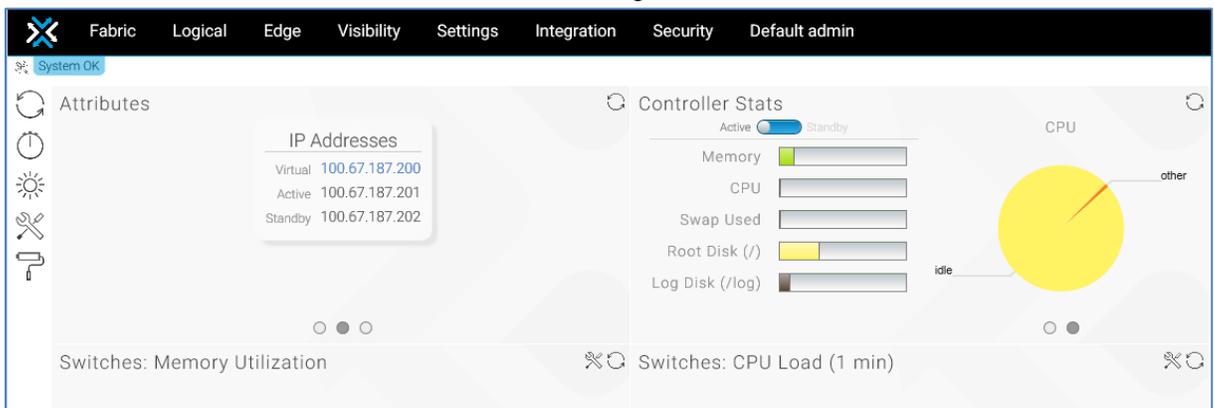


Figure 37 BCF dashboard

## 5.4 Switch deployment

### 5.4.1 Zero Touch Fabric overview

Big Switch Zero Touch Fabric (ZTF) uses the Open Networking Install Environment (ONIE) boot loader to automate switch installation and configuration. ONIE makes deploying many switches in a data center easier and less prone to errors. The ZTF process uses ONIE to automatically install the correct version of Switch Light OS on each switch when the switch is powered on and connected to the BCF Controller.

The Dell EMC Networking switches used in this example do not have BCF Switch Light OS installed initially. In the following steps, the BCF Controller deploys the OS to the leaf-spine fabric switches.

Switch Light OS is a complete SDN operating system based on Open Network Linux (ONL) and is bundled with the BCF software distribution. This ensures that the software running on the switch is compatible with the controller software version. Figure 38 shows the switch registration and OS deployment steps.

**Note:** For more information about this process, see Chapter 4 of the [Big Cloud Fabric 4.2.0 User Guide](#).

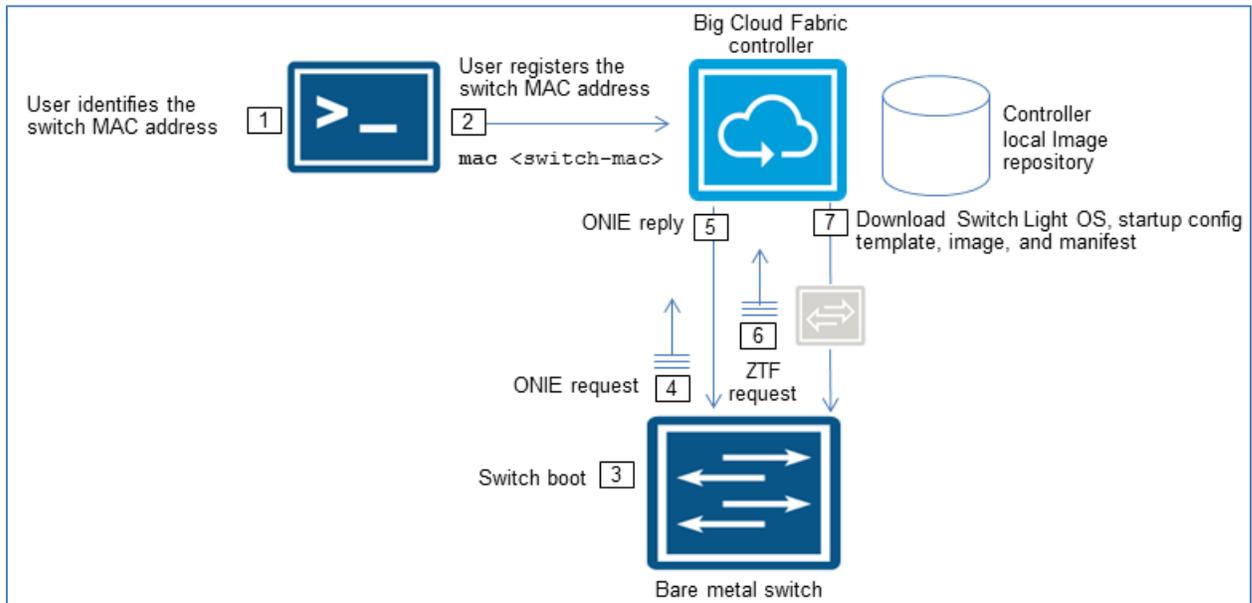


Figure 38 BCF switch registration and OS deployment workflow

A summary of the switch registration and OS deployment steps is provided in Table 3.

Table 3 BCF switch provisioning summary

| Step | Description   |
|------|---|
| 1    | Collect switch MAC address from the Dell EMC express service tag on the switch  |
| 2    | Register switch MAC address using the BCF GUI or CLI  |
| 3    | Switch is rebooted or power cycled  |
| 4    | The switch ONIE loader generates an IPv6 neighbor discovery message on the local network segment  |
| 5    | If the MAC is registered, the controller responds to the ONIE request from the switch and instructs it to download the Switch Light OS loader to begin installation |
| 6    | After installing the Switch Light OS loader and rebooting, the loader broadcasts a ZTF request  |
| 7    | The ZTF server on the active BCF Controller sends the Switch Light OS image, manifest, and startup configuration to the switch                                      |

The startup configuration from the controller includes the following information:

- Hostname
- Switch MAC address
- Controller IPv6 addresses
- NTP, logging, and Simple Network Management Protocol (SNMP) configuration

### 5.4.2 Collect switch MAC addresses

Record the MAC address of each leaf and spine switch. The MAC address is printed on the plastic express service tag labeled “EST” on each switch. The tag is located on the front of Z9100-ON switches, and the back of S4048-ON switches as shown in Figure 39.



Figure 39 EST tag location on Z9100-ON and S4048-ON switches

### 5.4.3 Provision switches in the BCF Controller

Table 4 lists the MAC addresses, switch names, roles, and leaf groups used for provisioning in this section.

Table 4 Switch provisioning details

| Model    | MAC address       | Switch name | Fabric role | Leaf group |
|----------|-------------------|-------------|-------------|------------|
| S4048-ON | f4:8e:38:20:37:29 | Leaf1       | Leaf        | Rack1      |
| S4048-ON | f4:8e:38:20:54:29 | Leaf2       | Leaf        | Rack1      |
| S4048-ON | 64:00:6a:e4:cc:3e | Leaf3       | Leaf        | Rack2      |
| S4048-ON | 64:00:6a:e7:24:14 | Leaf4       | Leaf        | Rack2      |
| Z9100-ON | 4c:76:25:e7:41:40 | Spine1      | Spine       | NA         |
| Z9100-ON | 4c:76:25:e7:3b:40 | Spine2      | Spine       | NA         |

The BCF Controller CLI or GUI may be used to provision the switches. The GUI is used in this example.

1. Using a browser, navigate to the VIP address of the BCF Controller cluster and log in.
2. Navigate to **Fabric > Switches** and click the + icon to open the **Provision Switch** dialog box.
3. To provision the leaf switches:
  - a. In the **Provision Switch** dialog box, enter or paste the **MAC address** of the first leaf switch as shown in Figure 40.

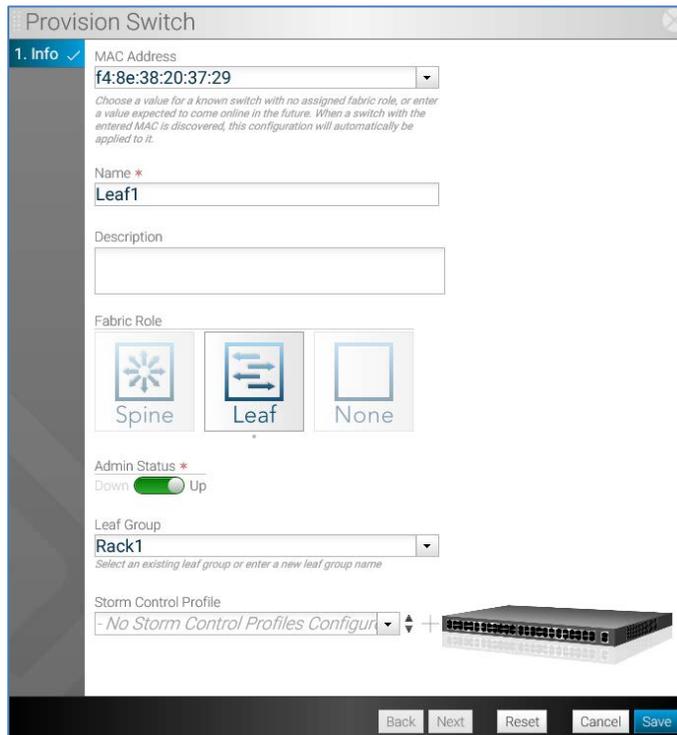


Figure 40 Provision the first leaf switch

- a. Enter the switch **Name**, select the **Fabric Role**, and enter the **Leaf Group** as listed in Table 4.
  - b. Make sure the **Admin Status** bar is green and set to **Up**.
  - c. Click **Save** and repeat for the remaining leaf switches.
4. To provision the spine switches, on the **Fabric > Switches** page, click the **+** icon to open the **Provision Switch** dialog box:
- a. In the **Provision Switch** dialog box, enter the **MAC address** of the first spine switch as shown in Figure 41.

Figure 41 Provision the first spine switch

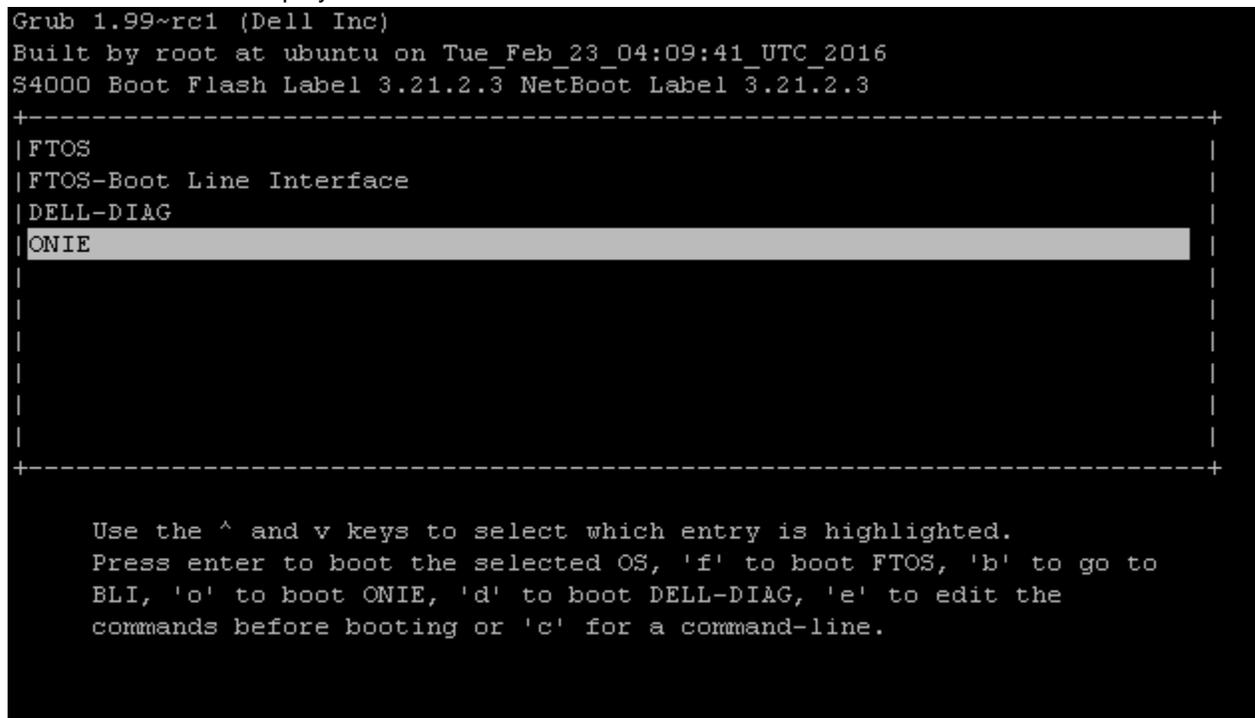
- a. Enter the switch **Name** and select the **Fabric Role** as listed in Table 4.
- b. Make sure the **Admin Status** bar is green and set to **Up**.
- c. Click **Save** and repeat for the remaining spine switch.

## 5.4.4 Boot switches in ONIE install mode

Before proceeding, make sure all leaf switches, spines switches, and BCF Controllers are physically connected to the p-switch network as shown in Figure 17 and Figure 18.

To place switches in ONIE install mode, do the following on each leaf and spine switch:

1. Power on or reboot the switch.
2. If **press Esc to stop autoboot** is shown during boot, press Esc. This step is required if switches are running the Dell Networking Operating System (DNOS) 9.x.
3. The Grub menu displays.



```
Grub 1.99~rc1 (Dell Inc)
Built by root at ubuntu on Tue_Feb_23_04:09:41_UTC_2016
S4000 Boot Flash Label 3.21.2.3 NetBoot Label 3.21.2.3
+-----+
| FTOS
| FTOS-Boot Line Interface
| DELL-DIAG
| ONIE
+-----+

Use the ^ and v keys to select which entry is highlighted.
Press enter to boot the selected OS, 'f' to boot FTOS, 'b' to go to
BLI, 'o' to boot ONIE, 'd' to boot DELL-DIAG, 'e' to edit the
commands before booting or 'c' for a command-line.
```

Figure 42 Grub menu on S4048-ON switch

4. In the Grub menu, select **ONIE** and press **Enter**.
5. In the next window, select **ONIE: Install OS** and press **Enter**.

This starts the Switch Light OS installation and configuration process, listed in steps 4-7 of Table 3. Allow a few minutes for the BCF Controller to install the Switch Light OS to each switch.

Optionally, provisioning progress may be monitored at the switch consoles. When provisioning is complete, the console of each switch appears with its hostname and login prompt as shown in Figure 43.

```
Switch Light OS SWL-OS-BCF-4.2.3 (0), 2017-08-26.00:51-ac47376
Leaf1 login: █
```

Figure 43 Leaf 1 console view after provisioning

### 5.4.5 Verify Switch Light OS installation

To verify successful installation of the Switch Light OS, use the BCF Controller GUI to navigate to **Fabric > Switches** to view all switches, MAC addresses, names, connection status, fabric status, and fabric roles.

Switches

Summary of Firmware Versions

| Loader Versions  | CPLD Versions  | ONIE Versions  |
|--|--|--|
| <a href="#">SWL-OS-BCF-4.2.3(0),2017-08-26.00:51-ac47376</a> 6 | <a href="#">6.4.4.4</a> 2<br><a href="#">15.12.5</a> 4 | <a href="#">3.21.1.2</a> 4<br><a href="#">3.23.1.3</a> 2 |

IP Address Allocation

Status ✗ Disabled

DNS Server —

Gateway —

Total Allocated Addresses 0

| Starting IP  | Ending IP | Subnet Mask Length | Addresses Allocated | Addresses Used | Utilization |
|--------------|-----------|--------------------|---------------------|----------------|-------------|
| No IP ranges |           |                    |                     |                |             |

| MAC               | Name                   | Description | Connected | Fabric Status  | Fabric Role | Spine | Leaf | Virtual | Leaf Group |
|-------------------|------------------------|-------------|-----------|--|-------------|-------|------|---------|------------|
| f4:8e:38:20:37:29 | <a href="#">Leaf1</a>  | —           | ✓         | ✓  | Leaf        | —     | ✓    | —       | Rack1      |
| f4:8e:38:20:54:29 | <a href="#">Leaf2</a>  | —           | ✓         | ✓  | Leaf        | —     | ✓    | —       | Rack1      |
| 64:00:6a:e4:cc:3e | <a href="#">Leaf3</a>  | —           | ✓         | ✓  | Leaf        | —     | ✓    | —       | Rack2      |
| 64:00:6a:e7:24:14 | <a href="#">Leaf4</a>  | —           | ✓         | ✓  | Leaf        | —     | ✓    | —       | Rack2      |
| 4c:76:25:e7:41:40 | <a href="#">Spine1</a> | —           | ✓         | ⚠ ASIC supported as spine only in forwarding-mode high-bandwidth or high-bandwidth-spine | Spine       | ✓     | —    | —       | NA         |
| 4c:76:25:e7:3b:40 | <a href="#">Spine2</a> | —           | ✓         | ⚠ ASIC supported as spine only in forwarding-mode high-bandwidth or high-bandwidth-spine | Spine       | ✓     | —    | —       | NA         |

Figure 44 Switch summary in BCF Controller GUI

## 5.5 Resolve common warnings and errors

Current warnings and errors may be viewed in the BCF Controller GUI by going to **Visibility > Fabric Summary**, or by clicking on the errors/warnings message in the upper left corner of the GUI.

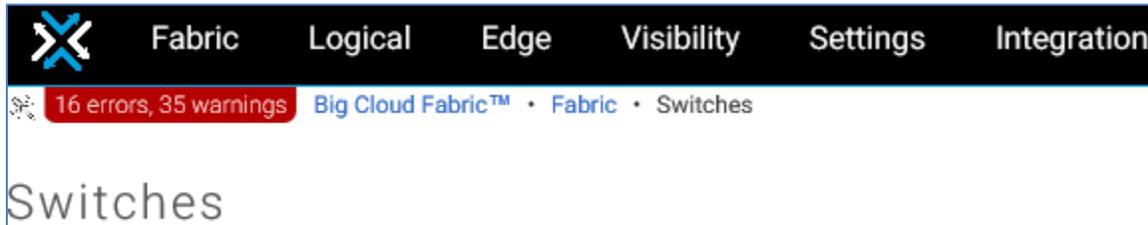
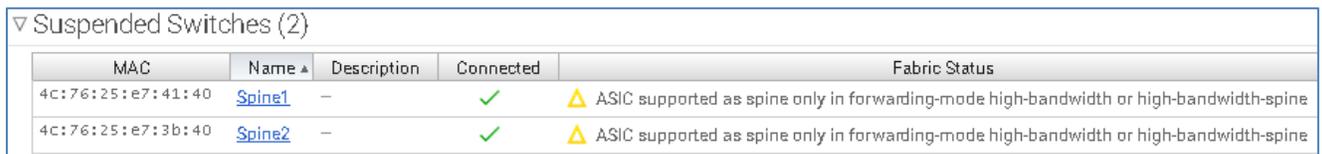


Figure 45 Errors/warnings message in GUI

### 5.5.1 Suspended Switches

Z9100-ON spine switches may appear under **Suspended Switches** with the message **ASIC supported as spine only in forwarding-mode high-bandwidth or high-bandwidth-spine**.



| MAC               | Name   | Description | Connected | Fabric Status  |
|-------------------|--------|-------------|-----------|--|
| 4c:76:25:e7:41:40 | Spine1 | -           | ✓         | ⚠ ASIC supported as spine only in forwarding-mode high-bandwidth or high-bandwidth-spine |
| 4c:76:25:e7:3b:40 | Spine2 | -           | ✓         | ⚠ ASIC supported as spine only in forwarding-mode high-bandwidth or high-bandwidth-spine |

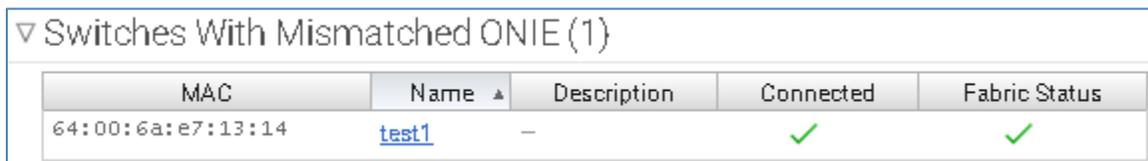
Figure 46 Suspended switches error

Z9100-ON switches are classified as high bandwidth spines in BCF. To set the forwarding mode to high bandwidth spine for these two switches, do the following:

1. Go to **Settings > Fabric Settings** and select the  icon.
2. In the left pane of the **Fabric Settings** dialog box, select **Forwarding Mode**.
3. In the right pane, move the **High Bandwidth Spine** slider to the right. All other sliders are moved to the left.
4. Click **Submit**.
5. Return to the **Visibility > Fabric Summary** page and verify there are no suspended switches listed.

### 5.5.2 Switches with mismatched ONIE and CPLD

Some switches may be listed with mismatched ONIE and/or Complex Programmable Logic Device (CPLD) firmware as shown in Figure 47 and Figure 48.



| MAC               | Name  | Description | Connected | Fabric Status |
|-------------------|-------|-------------|-----------|---------------|
| 64:00:6a:e7:13:14 | test1 | -           | ✓         | ✓             |

Figure 47 Switch with mismatched ONIE

| ▽ Switches With Mismatched CPLD (1) |                       |             |           |               |
|-------------------------------------|-----------------------|-------------|-----------|---------------|
| MAC                                 | Name ▲                | Description | Connected | Fabric Status |
| 64:00:6a:e7:13:14                   | <a href="#">test1</a> | -           | ✓         | ✓             |

Figure 48 Switch with mismatched CPLD

Resolve switch ONIE mismatches as follows:

1. Scroll down to **Switches With Mismatched ONIE** and click on the switch name. This example uses a single switch named **test1**.
2. In the switch page that opens, select the **Actions** tab.
3. On the left side of the page, select  **Manage Firmware**. The **Manage Switch Firmware** dialog box displays.

Manage Switch Firmware ✕

Switch: test1

| Firmware | Upgrade                                 | Current Version                              | Next Version                                 |
|----------|---|--|--|
| CPLD     | N <input checked="" type="checkbox"/> Y | 11.9.4                                       | 15.12.5                                      |
| Loader   | N <input checked="" type="checkbox"/> Y | SWL-OS-BCF-4.2.3(0),2017-08-26.00:51-ac47376 | SWL-OS-BCF-4.2.3(0),2017-08-26.00:51-ac47376 |
| ONIE     | N <input checked="" type="checkbox"/> Y | -  | 3.21.1.2                                     |

*CPLD and ONIE must be upgraded separately*

Reboot switch right away to effect upgrades

Figure 49 Manage switch firmware dialog box

4. Move the **CPLD** slider to **N**, and the **ONIE** slider to **Y**.

**Note:** CPLD and ONIE must be upgraded separately. Upgrade ONIE first.

5. Check the **Reboot switch right away** box and click **Upgrade**.
6. The switch reboots and ONIE firmware is updated. This can be observed at the switch console.
7. Repeat for remaining switches listed in the **Switches With Mismatched ONIE** table.
8. Refresh the **Visibility > Fabric Summary** page by clicking the  icon to verify all ONIE issues are resolved.

After ONIE mismatches are resolved, resolve switch CPLD mismatches as follows:

1. Scroll down to **Switches With Mismatched CPLD** and click on the switch name.
2. In the switch page that opens, select the **Actions** tab.
3. On the left side of the page, select  **Manage Firmware** to open the **Manage Switch Firmware** dialog box shown in Figure 49.
4. Ensure the **CPLD** slider is set to **Y**, and that the other sliders are set to **N**. Check the **Reboot switch right away** box and click **Upgrade**.
5. The switch reboots and CPLD firmware is updated. This can be observed at the switch console.
6. After the switch has rebooted and CPLD firmware installation is complete, power cycle the switch by removing the power cable(s), waiting until all LEDs are off (5-10 seconds), then reconnecting the power cable(s).
7. Repeat for remaining switches listed in the **Switches With Mismatched CPLD** table.
8. Refresh the **Visibility > Fabric Summary** page by clicking the  icon to verify all CPLD issues are resolved.

### 5.5.3 Switches without management address

Switches communicate with the controller using IPv6 on the p-switch control network. IPv6 addresses are automatically assigned to the fabric switches by the controller, but IPv4 management addresses are required if switches will connect to services that are not configured for IPv6 such as NTP, syslog, and SNMP.

**Note:** Switches connect to NTP, syslog, and SNMP servers using IPv4 addresses via the p-switch network. These connections are optional. See the [Big Cloud Fabric 4.2.0 User Guide](#) for more information.

The BCF Controller automatically assigns IPv4 management addresses from a defined address pool. Until this pool is configured, **Switches Without Management Addresses** are listed under **Errors** on the **Fabric Summary** page as shown in Figure 50.

| ▽ Switches Without Management Address (6) |                        |             |           |               |
|---|------------------------|-------------|-----------|---------------|
| MAC                                       | Name ▲                 | Description | Connected | Fabric Status |
| f4:8e:38:20:37:29                         | <a href="#">Leaf1</a>  | —           | ✓         | ✓             |
| f4:8e:38:20:54:29                         | <a href="#">Leaf2</a>  | —           | ✓         | ✓             |
| 64:00:6a:e4:cc:3e                         | <a href="#">Leaf3</a>  | —           | ✓         | ✓             |
| 64:00:6a:e7:24:14                         | <a href="#">Leaf4</a>  | —           | ✓         | ✓             |
| 4c:76:25:e7:41:40                         | <a href="#">Spine1</a> | —           | ✓         | ✓             |
| 4c:76:25:e7:3b:40                         | <a href="#">Spine2</a> | —           | ✓         | ✓             |

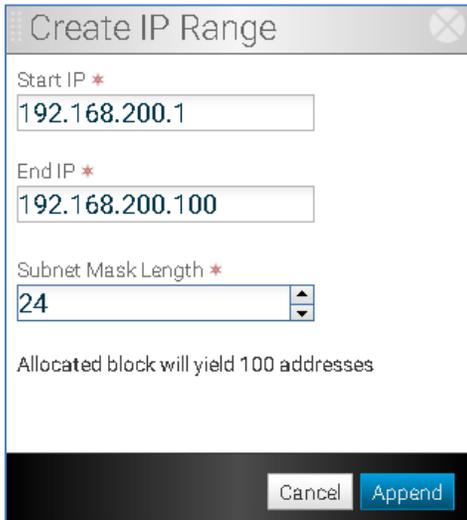
Figure 50 Switches without management address

The IPv4 address pool is configured as follows:

1. In the BCF GUI, go to **Fabric > Switches**.
2. Next to **IP Address Allocation**, click the  icon.
3. In the **Configure Switch IP Allocation** dialog box, move the slider to **Enabled**.
4. Click the  icon to open the **Create IP Range** dialog box.

**Note:** The **DNS Server Address** and **Gateway Address** fields are optional and not used in this example.

5. Specify a **Start IP**, **End IP**, and **Subnet Mask Length** to use for the pool. This example uses the range 192.168.200.1-100 with a subnet mask length of 24 as shown in Figure 51.



Create IP Range

Start IP \*  
192.168.200.1

End IP \*  
192.168.200.100

Subnet Mask Length \*  
24

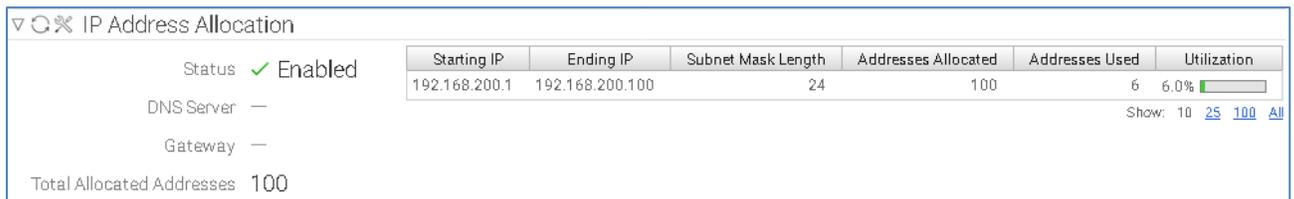
Allocated block will yield 100 addresses

Cancel Append

Figure 51 IP address range

6. Click **Append > Submit**.

When complete, the **IP Address Allocation** section of the **Fabric > Switches** page displays similar to Figure 52.



| Status    | Starting IP   | Ending IP       | Subnet Mask Length | Addresses Allocated | Addresses Used | Utilization |
|-----------|---------------|-----------------|--------------------|---------------------|----------------|-------------|
| ✓ Enabled | 192.168.200.1 | 192.168.200.100 | 24                 | 100                 | 6              | 6.0%        |

DNS Server —

Gateway —

Total Allocated Addresses 100

Show: 10 25 100 All

Figure 52 IP address pool configured

Notice that six addresses are used, one for each leaf and spine switch in the topology.

This resolves the **Switches Without Management Addresses** errors listed on the **Visibility > Fabric Summary** page.

## 5.5.4 Leaf interfaces not in interface groups

MLAGs are also referred to as interface groups in BCF. Connected leaf edge ports that are not in MLAGs display in the **Leaf Interfaces Not in Interface Groups** section. The VMware integration process covered in Section 7 automatically configures the interface groups and resolves these warnings.

 **Warnings**

▽ Leaf Interfaces Not in Interface Groups (24)

| Switch ▲              | Switch MAC        | Interface Name | Description | Status | Interface Group |
|-----------------------|-------------------|----------------|-------------|--------|-----------------|
| <a href="#">Leaf1</a> | f4:8e:38:20:37:29 | ethernet1      | —           | ✓ Up   | —               |
| <a href="#">Leaf1</a> | f4:8e:38:20:37:29 | ethernet2      | —           | ✓ Up   | —               |
| <a href="#">Leaf1</a> | f4:8e:38:20:37:29 | ethernet3      | —           | ✓ Up   | —               |
| <a href="#">Leaf1</a> | f4:8e:38:20:37:29 | ethernet4      | —           | ✓ Up   | —               |
| <a href="#">Leaf2</a> | f4:8e:38:20:54:29 | ethernet1      | —           | ✓ Up   | —               |
| <a href="#">Leaf2</a> | f4:8e:38:20:54:29 | ethernet2      | —           | ✓ Up   | —               |
| <a href="#">Leaf2</a> | f4:8e:38:20:54:29 | ethernet3      | —           | ✓ Up   | —               |
| <a href="#">Leaf2</a> | f4:8e:38:20:54:29 | ethernet4      | —           | ✓ Up   | —               |
| <a href="#">Leaf3</a> | 64:00:6a:e4:cc:3e | ethernet1      | —           | ✓ Up   | —               |
| <a href="#">Leaf3</a> | 64:00:6a:e4:cc:3e | ethernet2      | —           | ✓ Up   | —               |
| <a href="#">Leaf3</a> | 64:00:6a:e4:cc:3e | ethernet3      | —           | ✓ Up   | —               |
| <a href="#">Leaf3</a> | 64:00:6a:e4:cc:3e | ethernet4      | —           | ✓ Up   | —               |
| <a href="#">Leaf3</a> | 64:00:6a:e4:cc:3e | ethernet5      | —           | ✓ Up   | —               |
| <a href="#">Leaf3</a> | 64:00:6a:e4:cc:3e | ethernet6      | —           | ✓ Up   | —               |
| <a href="#">Leaf3</a> | 64:00:6a:e4:cc:3e | ethernet7      | —           | ✓ Up   | —               |
| <a href="#">Leaf3</a> | 64:00:6a:e4:cc:3e | ethernet8      | —           | ✓ Up   | —               |

Figure 53 Leaf interfaces not in interface groups

## 5.6 BCF validation commands from the CLI

The following commands help validate the fabric configuration. Run these commands from the active or standby controller.

**Note:** See the [Big Cloud Fabric 4.2.0 CLI Reference Guide](#) for a complete listing of commands.

### 5.6.1 show fabric error

The `show fabric error` command displays fabric errors. These items also appear in the GUI on the **Visibility > Fabric Summary** page under **Errors**. This command should return **None** at this point as shown below.

```
bcfctrl01> show fabric error
None.
```

**Note:** To see items shown in the GUI on the **Visibility > Fabric Summary** page under **Warnings**, run the command `show fabric warning`. At this stage of deployment, there are warnings shown for interfaces not configured in interface groups (MLAGs) as shown in section 5.5.4.

### 5.6.2 show link

The `show link` command returns all inter-switch links that are operational. This includes leaf-to-leaf (peer links) and leaf-spine links. Links are discovered using Link Layer Discovery Protocol (LLDP).

For the topology used in this deployment, shown in Figure 11, there are twelve inter-switch links: four peer links and eight leaf-spine links. All twelve inter-switch links should appear in the output as shown below:

```
bcfctrl01> show link
#  Switch Name  IF Name      Switch Name  IF Name      Link Type
--|-----|-----|-----|-----|-----|
1  Leaf1        ethernet53   Leaf2        ethernet53   peer
2  Leaf1        ethernet54   Leaf2        ethernet54   peer
3  Leaf3        ethernet53   Leaf4        ethernet53   peer
4  Leaf3        ethernet54   Leaf4        ethernet54   peer
5  Spine1       ethernet1    Leaf1        ethernet49   leaf-spine
6  Spine1       ethernet2    Leaf2        ethernet49   leaf-spine
7  Spine1       ethernet3    Leaf3        ethernet49   leaf-spine
8  Spine1       ethernet4    Leaf4        ethernet49   leaf-spine
9  Spine2       ethernet1    Leaf1        ethernet50   leaf-spine
10 Spine2       ethernet2    Leaf2        ethernet50   leaf-spine
11 Spine2       ethernet3    Leaf3        ethernet50   leaf-spine
12 Spine2       ethernet4    Leaf4        ethernet50   leaf-spine
```

### 5.6.3 show switch *switch name* interface

The command `show switch switch name interface` is used to check operational status of switch interfaces. Like most commands, it is run from the controller console instead of a switch console. The switch name is specified in the command.

```
bcfctrl01> show switch Spine1 interface
```

| # | Switch | IF Name   | IF Type | Phy. State | Op. State | LACP State | Curr Features  |
|---|--------|-----------|---------|------------|-----------|------------|----------------|
| 1 | Spine1 | ethernet1 | leaf    | up         | up        | inactive   | fiber, 40gb-fd |
| 2 | Spine1 | ethernet2 | leaf    | up         | up        | inactive   | fiber, 40gb-fd |
| 3 | Spine1 | ethernet3 | leaf    | up         | up        | inactive   | fiber, 40gb-fd |
| 4 | Spine1 | ethernet4 | leaf    | up         | up        | inactive   | fiber, 40gb-fd |

With BCF, the Z9100-ON interfaces are automatically configured for 40GbE as shown when connected to 40GbE interfaces on S4048-ON leaf switches.

**Note:** The command output above is truncated; the remaining Spine1 interfaces are down.

## 6 VMware vSphere deployment

VMware vSphere is a critical component of the deployment of the SDDC. This section provides an overview of the vSphere settings used for this deployment, and design decisions follow guidance outlined in VVD 4.1. Big Switch Networks recommends certain vSphere settings for integration with BCF, and those are included in this section where applicable.

For detailed vSphere deployment instructions, refer to the VMware documentation listed in Appendix D.

### 6.1 vCenter server deployment and design

In this deployment example, two vCenter Server appliances are deployed:

- mgmtvc01.dell.local – supports the ESXi hosts that compose the Management pod/cluster
- compvc01.dell.local – supports the ESXi hosts that compose the Compute pod/cluster

Deploying two VMware vCenter servers provides administrative and failure isolation benefits. By dividing along an administrative boundary, a separate security policy can be applied to either of the vCenter servers to reflect administrative functions that would typically be completed by separate organizations. As a secondary benefit, capacity planning for compute workloads is simplified by removing the management workloads from consideration. With this configuration, maintenance becomes easier and the compute workloads remain available during management workload maintenance windows.

Each vCenter Server is deployed using the Linux-based vCenter Server Appliance (VCSA). The VCSA is a prepackaged VM that is easy to deploy and supports up to 2000 hosts or 35,000 VMs.

Each vCenter server deploys with an external Platform Services Controller (PSC) which may be replicated when configured in external mode. With each PSC joined to a single vCenter single sign-on domain, the controllers function as a cluster and provide authentication to all components.

vCenter servers are assigned static IP addresses and hostnames during installation and include a valid DNS registration with reverse lookup. Table 5 shows the configuration information for the two vCenter Servers and their associated PSCs.

Table 5 vCenter servers and PSCs

| vCenter server component | IP address     | FQDN                |
|--------------------------|----------------|---------------------|
| Management vCenter       | 100.67.187.171 | mgmtvc01.dell.local |
| Management PSC           | 100.67.187.170 | mgmtpsc.dell.local  |
| Compute vCenter          | 100.67.187.173 | compvc01.dell.local |
| Compute PSC              | 100.67.187.172 | comppsc.dell.local  |

A vCenter Server has multiple sizing options available for selection during the deployment process. In this example, vCenter mgmtvc01 is built using the small appliance size (up to 100 hosts/1000 VMs, while vCenter compvc01 is built using the medium appliance size (up to 400 hosts/4000 VMs). See the [vSphere Installation and Setup](#) guide for more information on sizing and vCenter deployment instructions.

**Note:** Both vCenter servers and PSCs are deployed to hosts in the Management cluster.

After vCenter servers and PSCs are deployed, create a datacenter under each vCenter Server and a cluster under each datacenter. Table 6 outlines the vCenter components and initial host membership used in this example.

Table 6 Initial VMware vCenter Server configuration

| VMware vCenter Server | Datacenter name | Cluster name | Hostname          |
|-----------------------|-----------------|--------------|-------------------|
| mgmtvc01.dell.local   | MgmtDatacenter  | Management   | mgmt01.dell.local |
| mgmtvc01.dell.local   | MgmtDatacenter  | Management   | mgmt02.dell.local |
| mgmtvc01.dell.local   | MgmtDatacenter  | Management   | mgmt03.dell.local |
| mgmtvc01.dell.local   | MgmtDatacenter  | Management   | mgmt04.dell.local |
| compvc01.dell.local   | CompDatacenter  | Compute      | comp01.dell.local |
| compvc01.dell.local   | CompDatacenter  | Compute      | comp02.dell.local |
| compvc01.dell.local   | CompDatacenter  | Compute      | comp03.dell.local |
| compvc01.dell.local   | CompDatacenter  | Compute      | comp04.dell.local |

BCF requires all ESXi hosts have unique hostnames and that the domain name field not be empty. BCF recommends hosts use fully qualified domain names (FQDNs) as shown in the hostname column of Table 6.

**Note:** There are four hosts in each of the Management and Compute clusters in this example as shown in Figure 11. VVD 4.1 recommends at least four hosts per vSAN cluster. This allows you to take an ESXi host offline for maintenance without impacting the overall vSAN cluster health.

Figure 54 shows the **Hosts and Clusters** tab in the **VMware Web Client Navigator** pane. The two vSphere vCenter servers (🌐 icons) are shown at the top of each tree. Beneath each vCenter is its associated data center, cluster, hosts, and VMs. The two vCenter VMs and two PSC VMs are located in the Management cluster.

Five additional VMs running Microsoft Windows Server 2016 guest operating systems have also been added to validate traffic for this deployment guide:

- Management cluster – includes VMs 10 and 11, on Data Network 1
- Compute cluster – includes VMs 20 and 21 on Data Network 2, and VM 30 on Data Network 3

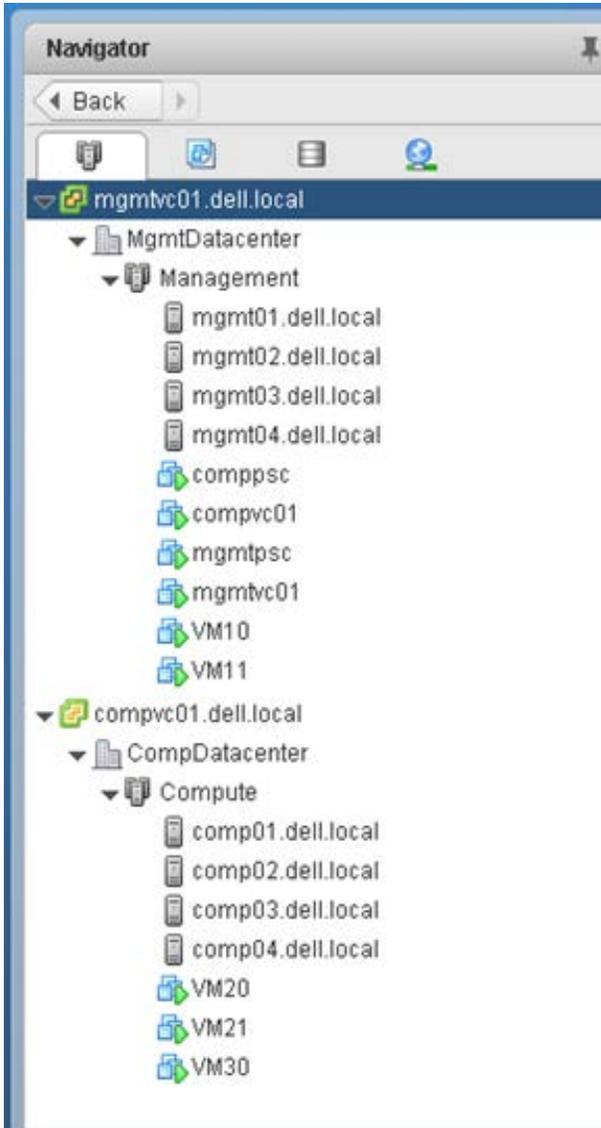


Figure 54 VMware Web Client - Hosts and Clusters tab

## 6.2 Virtual network design

When building the virtual network counterpart to BCF, a few principles are followed to ensure that the design meets a diverse set of requirements while keeping operational complexity to a minimum:

- The separation of network services to achieve greater security and performance by attaching each service to port groups with a different VLAN ID
- The use of network I/O control and traffic shaping that provides bandwidth to critical workloads
- All virtual machines use the VMXNET3 virtual NIC drivers

The VLANs and IP addresses used in this deployment are listed in Table 7. The L2/L3 boundary is at the leaf switches, and each VLAN is contained within a single rack.

Table 7 Production VLANs and IP addresses

| Cluster    | Function       | VLAN ID | Network        | Gateway       |
|------------|----------------|---------|----------------|---------------|
| Management | Data Network 1 | 1611    | 172.16.11.0/24 | 172.16.11.254 |
|            | vMotion        | 1612    | 172.16.12.0/24 | 172.16.12.254 |
|            | vSAN           | 1613    | 172.16.13.0/24 | None          |
| Compute    | Data Network 2 | 1621    | 172.16.21.0/24 | 172.16.21.254 |
|            | Data Network 3 | 1631    | 172.16.31.0/24 | 172.16.31.254 |
|            | vMotion        | 1622    | 172.16.22.0/24 | 172.16.22.254 |
|            | vSAN           | 1623    | 172.16.23.0/24 | None          |

The Data Network VLANs are used for VM-to-VM traffic. The number of Data Network VLANs used is entirely dependent on business needs. In this example, Data Network 1 is in the Management cluster, and Data Networks 2 and 3 are in the Compute cluster. In this deployment, gateways are configured to route traffic between the three Data Networks.

The vMotion VLANs also have gateways configured. This enables VMs to be migrated across racks as needed.

Each vSAN is on an isolated VLAN for best performance as recommended in VVD 4.1. No gateways are configured and vSAN traffic is contained within each rack.

The IP address and VLAN numbering scheme is similar to that used in VVD 4.1. Subnet-to-VLAN mapping uses the [RFC1918](#) defined private IP address space 172.16.0.0/12 as the base for all addresses. The second and third octets represent the VLAN ID.

For example, 172.16.11.0/24 has an associated VLAN ID of 1611. This algorithm ensures that each subnet and VLAN pairing is unique.

## 6.2.1 VDS configuration

This section provides details regarding VMware vSphere Distributed Switch (VDS) configuration. Following best practices, each VMware cluster (Management and Compute) has a single VDS to keep operational complexity to a minimum. For the Management cluster, the VDS is named VDS-Mgmt. For the Compute cluster, the VDS is named VDS-Comp.

In this example, the load balancing algorithm used for all port groups, regardless of the VDS, is IP Hash. IP Hash selects the uplink based on a hash of the source and destination IP address of each packet. This enables a single VM communicating with multiple IP addresses to load balance across multiple network adapters.

BCF automatically creates an MLAG on leaf switches with all physical links to the host when the VMware integration process is complete.

### 6.2.1.1 VDS-Mgmt configuration

Configuration settings used for VDS-Mgmt in the Management Cluster are listed in Table 8.

Table 8 VDS-Mgmt settings

| VDS switch name | Network I/O control | Number of uplinks | Discovery Protocol Type / Operation | MTU (Bytes) |
|-----------------|---------------------|-------------------|-------------------------------------|-------------|
| VDS-Mgmt        | Enabled             | 2                 | LLDP / Both                         | 9000 Bytes  |

**Note:** For **Discovery Protocol Type**, BCF supports CDP and LLDP. Since VDS-Mgmt includes vSAN and vMotion traffic, setting the MTU to its maximum value of 9000 is recommended for best performance.

The port group settings used for VDS-Mgmt are shown in Table 9.

Table 9 VDS-Mgmt port group settings

| Port group   | VLAN ID | Teaming and failover settings |                           |                 |          |                |
|--------------|---------|-------------------------------|---------------------------|-----------------|----------|----------------|
|              |         | Load balancing                | Network failure detection | Notify switches | Failback | Active uplinks |
| data1-mgmt   | 1611    | Route based on IP hash        | Link status only          | Yes             | Yes      | 1,2            |
| vmotion-mgmt | 1612    | Route based on IP hash        | Link status only          | Yes             | Yes      | 1,2            |
| vsan-mgmt    | 1613    | Route based on IP hash        | Link status only          | Yes             | Yes      | 1,2            |

**Note:** BCF recommends using **Route based on IP hash** as the load balancing method and does not support LACP configurations in vCenter.

### 6.2.1.2 VDS-Comp configuration details

Configuration settings used VDS-Comp in the Compute cluster are listed in Table 10.

Table 10 VDS-Comp settings

| VDS switch name | Network I/O control | Number of uplinks | Discovery Protocol Type / Operation | MTU (Bytes) |
|-----------------|---------------------|-------------------|-------------------------------------|-------------|
| VDS-Comp        | Enabled             | 4                 | LLDP / Both                         | 9000 Bytes  |

**Note:** For the discovery protocol type, BCF supports CDP and LLDP. Since VDS-Mgmt includes vSAN and vMotion traffic, setting the MTU to its maximum value of 9000 is recommended for best performance.

The port group settings used for VDS-Comp are shown in Table 11.

Table 11 VDS-Comp port group settings

| Port group   | VLAN ID | Teaming and failover settings |                           |                 |          |                |
|--------------|---------|-------------------------------|---------------------------|-----------------|----------|----------------|
|              |         | Load balancing                | Network failure detection | Notify switches | Failback | Active uplinks |
| data2-comp   | 1621    | Route based on IP hash        | Link status only          | Yes             | Yes      | 1,2,3,4        |
| data3-comp   | 1631    | Route based on IP hash        | Link status only          | Yes             | Yes      | 1,2,3,4        |
| vmotion-comp | 1622    | Route based on IP hash        | Link status only          | Yes             | Yes      | 1,2,3,4        |
| vsan-comp    | 1623    | Route based on IP hash        | Link status only          | Yes             | Yes      | 1,2,3,4        |

**Note:** BCF recommends using Route based on IP hash as the load balancing method and does not support LACP configurations in vCenter.

### 6.2.1.3 VDS summary

The **Networking** tab in the **VMware Web Client Navigator** pane is shown in Figure 55 after VDS configuration is complete. Each VDS (**VDS-Mgmt** and **VDS-Comp**) is shown with its configured port groups and uplinks.

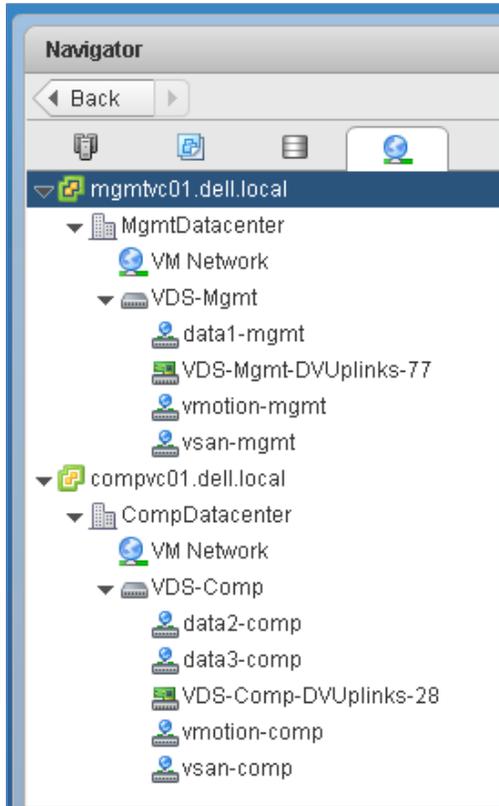


Figure 55 VMware Web Client - Networking tab

## 6.2.2 VMkernel adapter configuration

VMkernel adapters provide connectivity to hosts and handle ESXi management, vMotion, and vSAN traffic. In this section, VMkernel adapters are created and associated with VDS port groups.

During ESXi installation, vmk0 is automatically created on a VMware standard switch (VSS) named vSwitch0, for the OOB management network. No further configuration is needed for this connection. Two additional VMkernel adapters, vmk1 and vmk2, are manually added to each ESXi host for vMotion and vSAN traffic. These adapters are connected to the VDS in each cluster.

The vSAN VMkernel adapter is used for vSAN traffic within the cluster and uses the **Default** TCP/IP stack.

The vMotion VMkernel adapter is configured to allow VM mobility within and across clusters and it uses the **vMotion** TCP/IP stack. The vMotion TCP/IP stack allows a dedicated default gateway to be specified. This enables vMotion traffic to be routed between networks and racks.

### 6.2.2.1 Management cluster hosts

Table 12 shows the VMkernel configuration details for hosts in the Management cluster:

Table 12 VDS-Mgmt VMkernel adapters

| VDS      | Existing network | TCP/IP stack | Enabled services | Host VMkernel IP addresses | TCP/IP stack gateway address | MTU  |
|----------|------------------|--------------|------------------|----------------------------|------------------------------|------|
| VDS-Mgmt | vmotion-mgmt     | vMotion      | vMotion          | 172.16.12.1-4 /24          | 172.16.12.254                | 9000 |
| VDS-Mgmt | vsan-mgmt        | Default      | vSAN             | 172.16.13.1-4 /24          | Default (Not Used)           | 9000 |

When configuration is complete, the **VMkernel adapters** page for each host in the Management cluster appears similar to Figure 56. The vMotion adapter, selected, has a default gateway configured to enable vMotion traffic to be routed between racks.

**VMkernel adapters**

| Device | Network Label      | Switch   | IP Address    | TCP/IP Stack |
|--------|--------------------|----------|---------------|--------------|
| vmk0   | Management Netw... | vSwitch0 | 100.67.187.19 | Default      |
| vmk1   | vmotion-mgmt       | VDS-Mgmt | 172.16.12.1   | vMotion      |
| vmk2   | vsan-mgmt          | VDS-Mgmt | 172.16.13.1   | Default      |

**VMkernel network adapter: vmk1**

**All** Properties IP Settings Policies

**Port properties**

|                  |              |
|------------------|--------------|
| Network label    | vmotion-mgmt |
| TCP/IP stack     | vMotion      |
| Enabled services | vMotion      |

**IPv4 settings**

|                 |                      |
|-----------------|----------------------|
| DHCP            | Disabled             |
| IPv4 address    | 172.16.12.1 (static) |
| Subnet mask     | 255.255.255.0        |
| Default gateway | 172.16.12.254        |

Figure 56 VMkernel adapters for host mgmt01

Figure 57 shows the completed topology of VDS-Mgmt for the Management cluster showing port groups, VLAN assignments, VMkernels, IP addresses, and physical NIC uplinks.

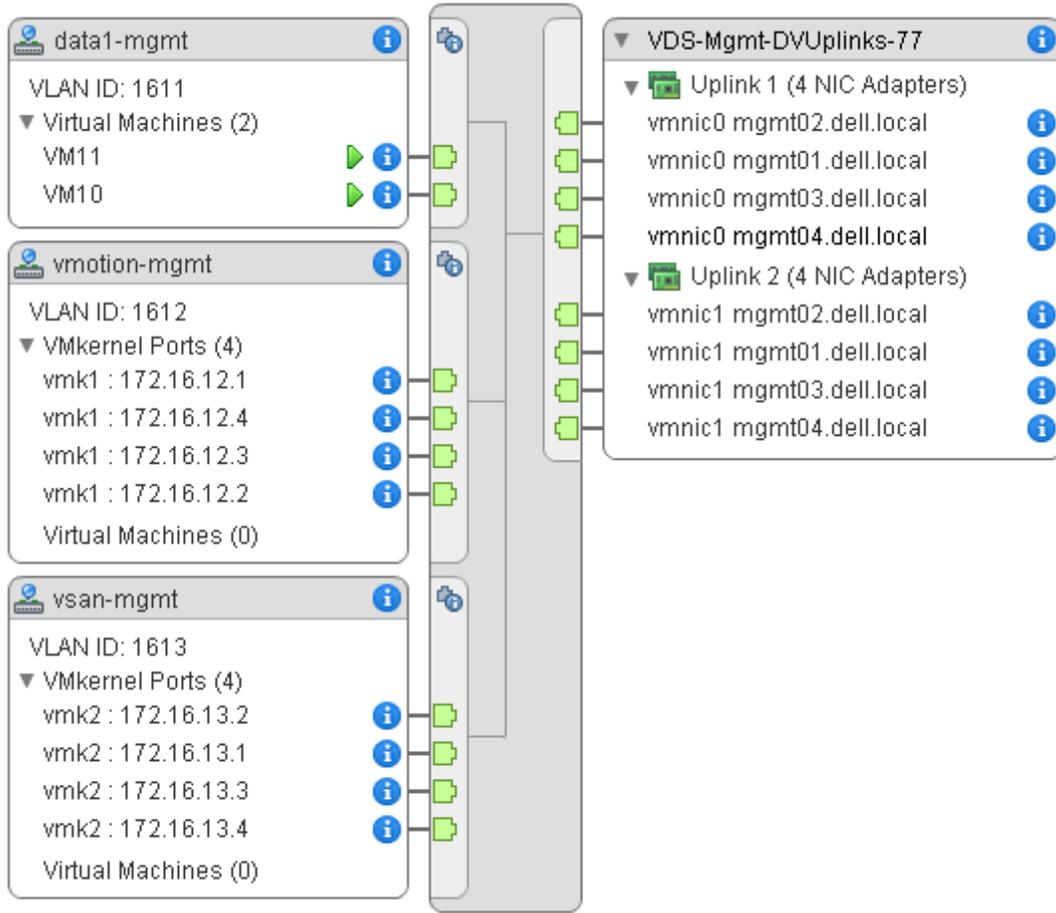


Figure 57 VDS-Mgmt topology

### 6.2.2.2 Compute cluster hosts

The VMkernel configuration details for hosts in the Compute cluster are listed in Table 13.

Table 13 VDS-Comp VMkernel adapters

| VDS      | Existing network | TCP/IP stack | Enabled services | Host VMkernel IP addresses | TCP/IP stack gateway address | MTU  |
|----------|------------------|--------------|------------------|----------------------------|------------------------------|------|
| VDS-Comp | vmotion-mgmt     | vMotion      | vMotion          | 172.16.22.1-4 /24          | 172.16.22.254                | 9000 |
| VDS-Comp | vsan-mgmt        | Default      | vSAN             | 172.16.23.1-4 /24          | Default (Not Used)           | 9000 |

When configuration is complete, the **VMkernel adapters** page for each host in the Compute cluster displays as shown in Figure 58. The vMotion adapter, selected, has a default gateway configured to enable vMotion traffic to be routed between racks.

**VMkernel adapters**

| Device | Network Label      | Switch   | IP Address    | TCP/IP Stack |
|--------|--------------------|----------|---------------|--------------|
| vmk0   | Management Network | vSwitch0 | 100.67.187.20 | Default      |
| vmk1   | vmotion-comp       | VDS-Comp | 172.16.22.1   | vMotion      |
| vmk2   | vsan-comp          | VDS-Comp | 172.16.23.1   | Default      |

**VMkernel network adapter: vmk 1**

**All** Properties IP Settings Policies

**Port properties**

|                  |              |
|------------------|--------------|
| Network label    | vmotion-comp |
| TCP/IP stack     | vMotion      |
| Enabled services | vMotion      |

**IPv4 settings**

|                 |                      |
|-----------------|----------------------|
| DHCP            | Disabled             |
| IPv4 address    | 172.16.22.1 (static) |
| Subnet mask     | 255.255.255.0        |
| Default gateway | 172.16.22.254        |

Figure 58 VMkernel adapters for host comp01

Figure 59 shows the completed topology of VDS-Comp for the Compute cluster showing port groups, VLAN assignments, VMkernel, IP addresses, and physical NIC uplinks.

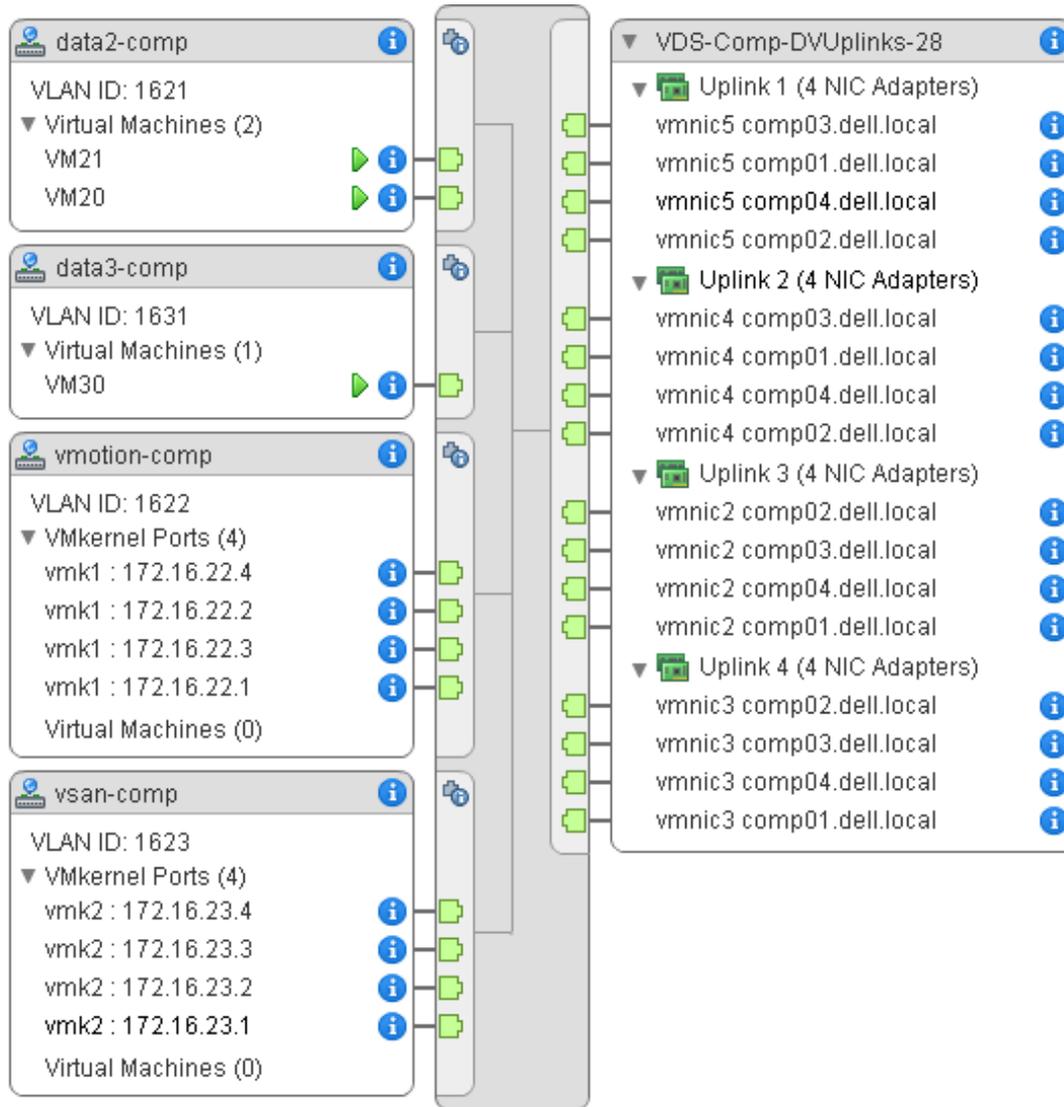


Figure 59 VDS-Comp topology

**Note:** Hosts and VMs will not be able to communicate with each other on the Data, vSAN and vMotion networks until the vCenters are integrated with BCF in the next section.

## 7 VMware integration with BCF

Integrating VMware vSphere with BCF provides an integrated solution that uses BCF as the underlying physical network. Integration benefits include:

- Automatic BCF ToR-to-host link detection and MLAG formation
- Automatic BCF L2 network creation and VM learning
- Network policy migration for vMotion / DRS
- Improved VM network visibility and troubleshooting, especially in regard to mapping between virtual and physical network resources

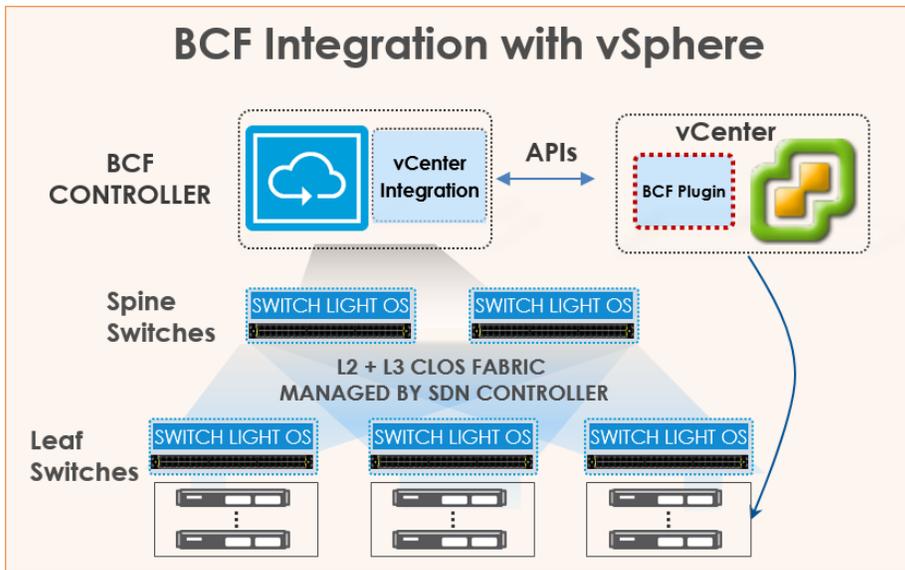


Figure 60 BCF integration with VMware vSphere

The information used to integrate both vCenter servers with BCF is shown in Table 14. With BCF automation set to **Full**, the BCF configuration is automatically updated in response to changes on vCenter.

**Note:** The other BCF automation levels are **On Demand** and **None**. Refer to the [BCF 4.2.0 User Guide](#) for more information.

The vCenter plugin access right sets the permission level for the vCenter BCF plugin. The **Read-Write** option allows the plugin to be used similarly to the BCF GUI from within vCenter. It may be set to **Read-Only** to prevent changes to BCF from the vCenter plugin.

Table 14 VMware vCenter connection details

| Name     | Hostname            | Tenant   | BCF configuration automation level | vCenter plugin access right |
|----------|---------------------|----------|------------------------------------|-----------------------------|
| mgmtvc01 | mgmtvc01.dell.local | mgmtvc01 | Full                               | Read-Write                  |
| compvc01 | compvc01.dell.local | compvc01 | Full                               | Read-Write                  |

## 7.1 Add vCenters to BCF

In this section, the Management and Compute vCenters are added to BCF and the automatic configuration level is set to full.

1. In the BCF GUI, navigate to **Integration > Orchestration > VMware vCenters**.
2. Select the **+** icon to open the **Create vCenter** dialog box.
3. In the **Create vCenter** dialog box, complete the following:
  - a. **Name:** Provide the name of the first vCenter, **mgmtvc01**.
  - b. **Hostname:** Enter the FQDN of the first vCenter, **mgmtvc01.dell.local**
  - c. **Username/Password:** Provide the vCenter login credentials
  - d. **Tenant:** Click the **+** icon to create a new tenant named **mgmtvc01**. Leave **Multicast** set to **Disabled** and click **Submit**.

**Note:** As of vSAN 6.6, vSAN communication is done via unicast and no multicast configuration is required.

- e. Leave **Operational Mode** set to **Normal** and **Preserve Auto-Generated BCF Configuration** set to **No**.
- f. In this example, the **vCenter Plugin Access right** is set to **Read-Write**.
- g. Set the **BCF Config Automation Level** to **Full**. This enables changes in vCenter to automatically update the BCF Controller without manual intervention. Figure 61 shows the configuration to this point.

**For complete integration ...**

- All ESXi hosts should have unique hostnames.
- CDP or LLDP should be enabled on vSphere virtual switches that connect to Big Cloud Fabric.

Name \*  
mgmtvc01

Host Name  
mgmtvc01.dell.local  
Up to 255 characters in length

Description

Username  
administrator@dell.local  
Up to 255 characters in length

Password  
●●●●●●  
 Show Password  
Up to 255 characters in length

Tenant  
mgmtvc01

Operational Mode  
Maintenance  Normal  
Set to **Maintenance Mode** to disconnect vCenter instance from BCF controller while preserving configuration.

Preserve Auto-Generated BCF Configuration  
No  Yes  
Set to **Yes** to preserve any auto-generated configurations when this vCenter instance is deleted. Choose **No** to force any auto-generated configurations to be cleaned up automatically when this vCenter instance is deleted.

vCenter Plugin Access Right \*  
Read-Write  
User will have permission to create, edit, and delete tenant, segments, logical interfaces, and routes.

BCF Config Automation Level \*  
Full  
Monitor vCenter and automatically configure BCF based on vCenter network configuration.

Back Next Reset Cancel Save

Figure 61 Create vCenter dialog box

4. Click **Next** to open the **Excluded VLANs** page.
  - a. Next to **VLAN Ranges to Exclude**, click the **+** icon.
  - b. Set the range from **0** to **0** to exclude untagged traffic.
5. Click **Save**.

Repeat the steps above for the second vCenter, **compvc01**.

The vCenters display on the **Integration > Orchestration > VMware vCenters** page as shown in Figure 62:

|                          | Name     | Operating Mode | Description | Status | Status Detail               | Hostname            | Username                 | vCenter Plugin Access Right | Tenant   | Version | Configuration Automation Level |
|--------------------------|----------|----------------|-------------|--------|-----------------------------|---------------------|--------------------------|-----------------------------|----------|---------|--------------------------------|
| <input type="checkbox"/> | compvc01 | Normal         | -           | ✓      | Connected and authenticated | compvc01.dell.local | administrator@dell.local | Read-Only                   | compvc01 | 6.5.0   | Full                           |
| <input type="checkbox"/> | mgmtvc01 | Normal         | -           | ✓      | Connected and authenticated | mgmtvc01.dell.local | administrator@dell.local | Read-Write                  | mgmtvc01 | 6.5.0   | Full                           |

Figure 62 VMware vCenters integrated with BCF

BCF imports the vCenter configuration and automatically configures the switches. Clicking the vCenter name provides configuration details for that vCenter.

For example, clicking **compvc01** displays the page shown in Figure 63. The **Info** and **Graphic** items are selected in the left pane. In the right pane, the **Info** section provides an overview of the configuration. Items in the **Graphic** section are selected to display information such as the mapping of vmnics to physical switch ports and associated port groups.

Figure 63 vCenter details for compvc01

Automatically configured interface groups, or MLAGs, are viewed in BCF by going to **Fabric > Interface Groups** as shown in Figure 64.

|                          | Name  | Description  | State  | Mode | Auto-Discovered | Backup Mode | Preempt Backups when Primaries Available | Total Member Interfaces | Member Interface Status | Total Backup Member Interfaces | Backup Member Interface Status | Total Host Interfaces | Leaf Group |
|--------------------------|---|--|--------|------|-----------------|-------------|--|-------------------------|-------------------------|--------------------------------|--------------------------------|-----------------------|------------|
| <input type="checkbox"/> | <a href="#">mgmt04.dell.local-VDS-Mgmt</a>        | Interface group for virtual switch VDS-Mgmt in ESXi host mgmt04.dell.local | ✓ Up   | LLDP | -               | Static      | -  | 2                       | ✓ All Up                | 0                              | NA                             | 2                     | Rack1      |
| <input type="checkbox"/> | <a href="#">mgmt03.dell.local-VDS-Mgmt</a>        | Interface group for virtual switch VDS-Mgmt in ESXi host mgmt03.dell.local | ✓ Up   | LLDP | -               | Static      | -  | 2                       | ✓ All Up                | 0                              | NA                             | 2                     | Rack1      |
| <input type="checkbox"/> | <a href="#">mgmt02.dell.local-VDS-Mgmt</a>        | Interface group for virtual switch VDS-Mgmt in ESXi host mgmt02.dell.local | ✓ Up   | LLDP | -               | Static      | -  | 2                       | ✓ All Up                | 0                              | NA                             | 2                     | Rack1      |
| <input type="checkbox"/> | <a href="#">mgmt01.dell.local-VDS-Mgmt</a>        | Interface group for virtual switch VDS-Mgmt in ESXi host mgmt01.dell.local | ✓ Up   | LLDP | -               | Static      | -  | 2                       | ✓ All Up                | 0                              | NA                             | 2                     | Rack1      |
| <input type="checkbox"/> | <a href="#">comp04.dell.local-VDS-Comp</a>        | Interface group for virtual switch VDS-Comp in ESXi host comp04.dell.local | ✓ Up   | LLDP | -               | Static      | -  | 4                       | ✓ All Up                | 0                              | NA                             | 4                     | Rack2      |
| <input type="checkbox"/> | <a href="#">comp03.dell.local-VDS-Comp</a>        | Interface group for virtual switch VDS-Comp in ESXi host comp03.dell.local | ✓ Up   | LLDP | -               | Static      | -  | 4                       | ✓ All Up                | 0                              | NA                             | 4                     | Rack2      |
| <input type="checkbox"/> | <a href="#">comp02.dell.local-VDS-Comp</a>        | Interface group for virtual switch VDS-Comp in ESXi host comp02.dell.local | ✓ Up   | LLDP | -               | Static      | -  | 4                       | ✓ All Up                | 0                              | NA                             | 4                     | Rack2      |
| <input type="checkbox"/> | <a href="#">comp01.dell.local-VDS-Comp</a>        | Interface group for virtual switch VDS-Comp in ESXi host comp01.dell.local | ✓ Up   | LLDP | -               | Static      | -  | 4                       | ✓ All Up                | 0                              | NA                             | 4                     | Rack2      |
| <input type="checkbox"/> | <a href="#">mgmt04.dell.local-vSwitch0-vmnic4</a> | Interface group for virtual switch vSwitch0 in ESXi host mgmt04.dell.local | ✗ Down | CDP  | -               | Static      | -  | 0                       | NA                      | 0                              | NA                             | 1                     | -          |

Figure 64 Interface Groups page

All host-to-leaf switch connections are listed and their status is **Up**. In this deployment, there are two member interfaces from each management host and four member interfaces from each compute host.

**Note:** BCF also reads OOB management network (vSwitch0) information from vCenter. The last row in Figure 64 shows the first of these connections as **Down** with **0** member interfaces. These connections are to the S3048-ON switches and are not managed by BCF. This information can be disregarded.

Interface group details are viewed by clicking the ▶ next to the hostname as shown in Figure 65:

Interface group for virtual switch VDS-Comp in ESXi host comp01.dell.local

4 ✓ All Up 0 NA

### Member Interfaces

|       | Switch            | Switch MAC | Interface Name | Description | Status | Spine Switch | Leaf Switch | Virtual Switch | Interface Group Member State |          |
|-------|-------------------|------------|----------------|-------------|--------|--------------|-------------|----------------|------------------------------|----------|
|       |                   |            |                |             |        |              |             |                | Operational                  | Physical |
| Leaf3 | 64:00:6a:e4:cc:3e | ethernet1  | -              | ✓ Up        | -      | ✓            | -           | ✓ Up           | ✓ Up                         |          |
| Leaf3 | 64:00:6a:e4:cc:3e | ethernet2  | -              | ✓ Up        | -      | ✓            | -           | ✓ Up           | ✓ Up                         |          |
| Leaf4 | 64:00:6a:e7:24:14 | ethernet1  | -              | ✓ Up        | -      | ✓            | -           | ✓ Up           | ✓ Up                         |          |
| Leaf4 | 64:00:6a:e7:24:14 | ethernet2  | -              | ✓ Up        | -      | ✓            | -           | ✓ Up           | ✓ Up                         |          |

### Backup Interfaces

No interfaces

### Host Interfaces

| Host Name         | Interface Name |
|-------------------|----------------|
| comp01.dell.local | vmnic2         |
| comp01.dell.local | vmnic3         |
| comp01.dell.local | vmnic4         |
| comp01.dell.local | vmnic5         |

Figure 65 Interface group details for comp01 in BCF

Information includes connection details to each leaf switch by port and host interface vmnic names.

## 7.2 Add BCF Plugin to vCenter

Adding the BCF plugin to vCenter is optional. The plugin lets you monitor and configure certain BCF components from the vSphere Web Client as an alternative to using the BCF GUI.

**Note:** For more information about the BCF plugin, see the [Big Cloud Fabric 4.2.0 User Guide](#).

The installation wizard is accessed in the BCF GUI by going to **Integration > Orchestration > VMware vCenters**. Select the ≡ icon next to the vCenter name and click **Deploy vCenter GUI Plugin**.

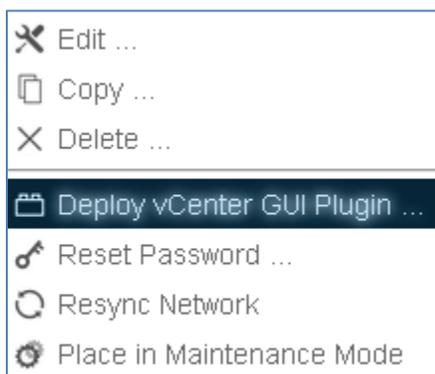


Figure 66 Deploy BCF plugin to vCenter

Enter the vCenter username and password in the dialog box and click **Submit**.

After installation is complete, log out and log back in to the vSphere Web Client for the vCenter. The **Home** page displays the **Big Cloud Fabric** icon.

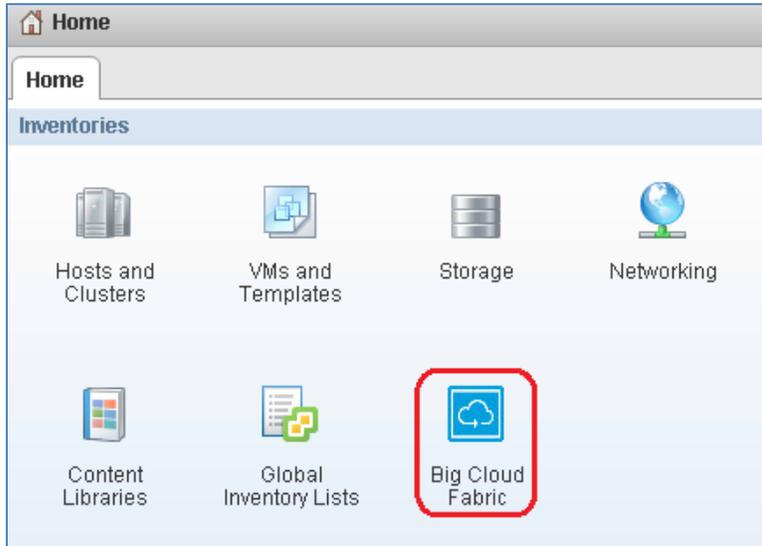


Figure 67 Big Cloud Fabric icon in vSphere Web Client

In the vSphere Web Client, double-click on the **Big Cloud Fabric** icon to open the page. Click on the **BCF Pod** address, **100.67.187.200**. The **Overview** page displays.

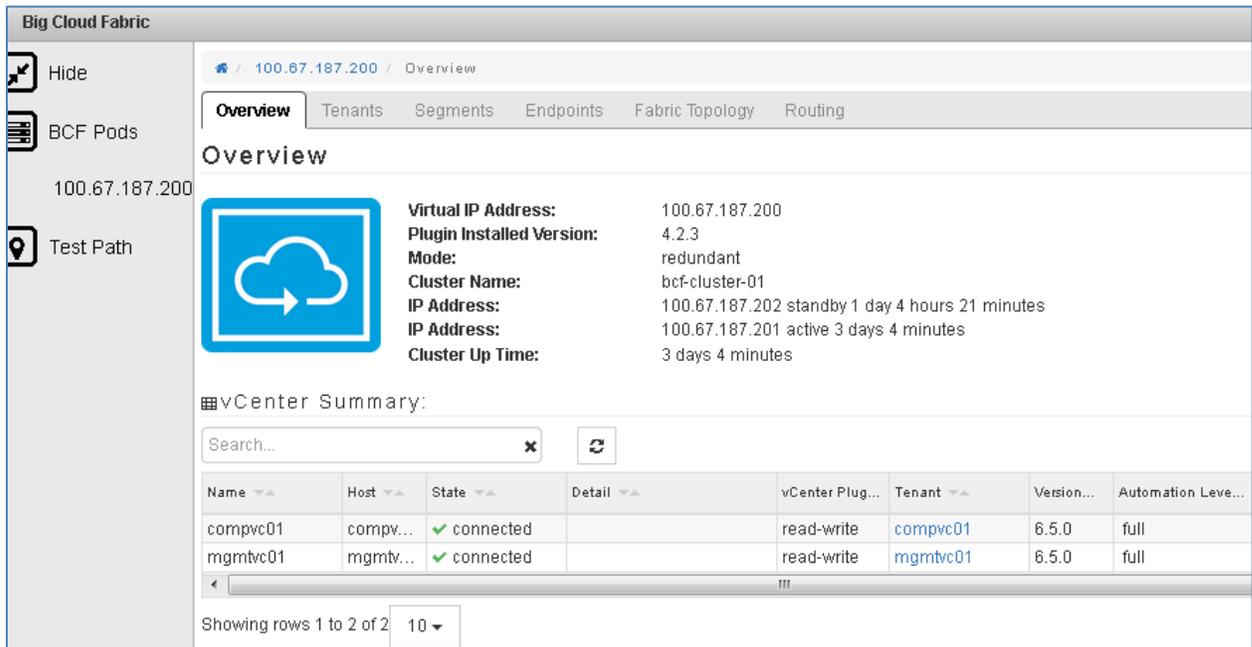


Figure 68 BCF plugin for vSphere web client

# 8 BCF tenant and segment configuration

## 8.1 Overview

In BCF, a tenant provides a logical grouping of layer 2 and layer 3 networks and services and is similar in function to a Virtual Routing and Forwarding (VRF) entity. Each tenant establishes a layer 3 boundary that separates traffic from other tenants through a logical router.

A segment is a layer 2 network, which consists of logical ports and endpoints. Within each tenant, separate segments establish layer 2 boundaries for each tier. VMkernel adapters and VM vNICs are endpoints.

Figure 69 shows the two vCenter tenants in this deployment, mgmtvc01 and compvc01, and their respective segments, hosts, and VMs. The BCF system tenant is built in and configured as needed to pass traffic between tenants.

**Note:** Only two of the four hosts from each cluster are shown in Figure 69 for clarity.

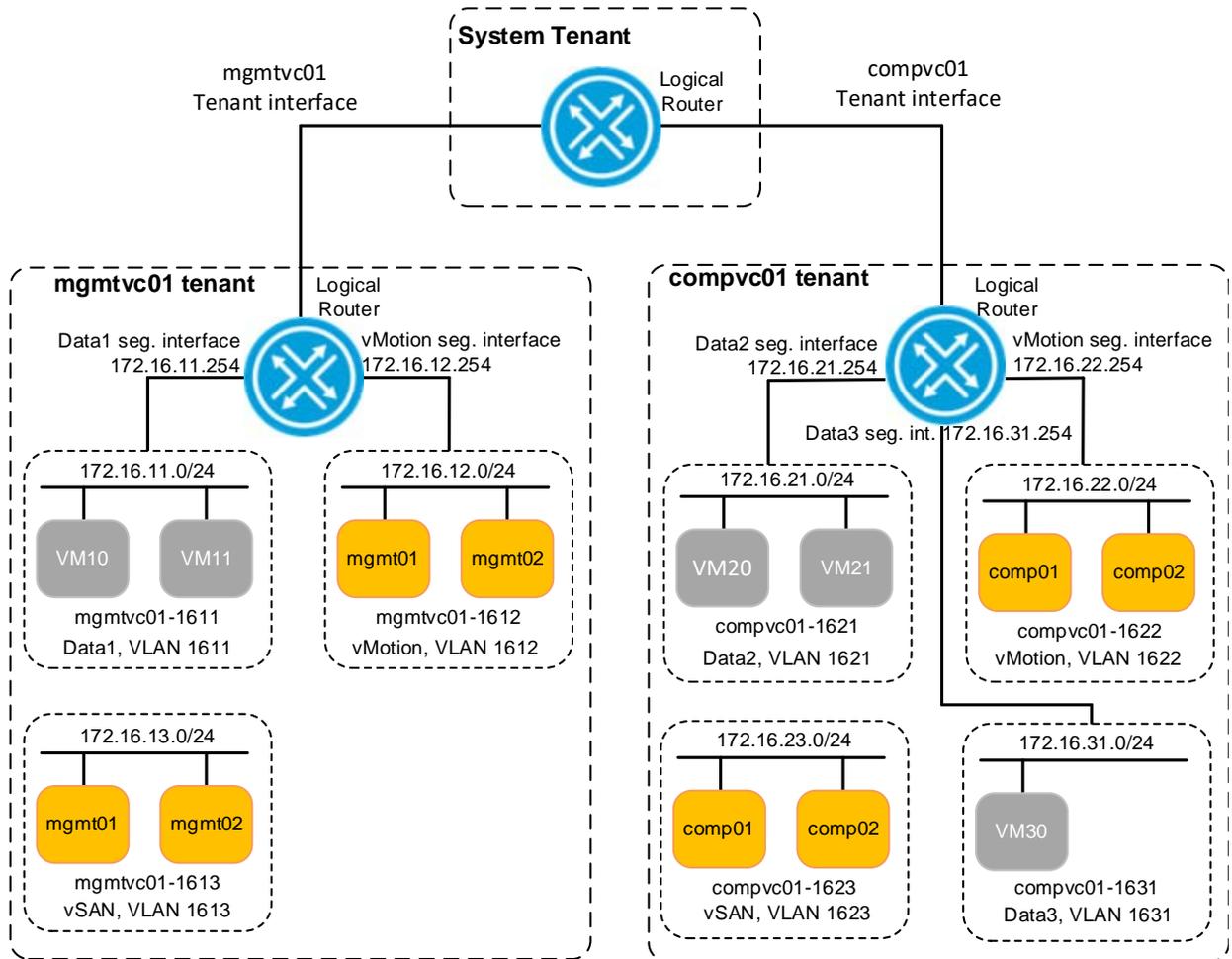


Figure 69 BCF tenant and segments

A logical router is automatically assigned to each tenant when it is defined. The logical router has two types of interfaces: tenant interfaces, and segment interfaces. Tenant interfaces are used to connect tenants together via the built in system tenant. Segment interfaces act as gateways for forwarding between segments within a tenant and routing traffic to other tenants through the system tenant.

In this chapter, tenant and segment interfaces are configured to enable communication between segments and tenants.

## 8.2 View tenants and segments

One tenant is manually created for each vCenter during the VMware integration process (completed in section 7.1). The two vCenter tenants are visible in the BCF GUI by selecting **Logical > Tenants** as shown in Figure 70.

| Name     | Description            | Multicast Enabled | Router MAC Address | Applied Policy List | System Tenant Interface | BGP Configured | vCenter  |
|----------|------------------------|-------------------|--------------------|---------------------|-------------------------|----------------|----------|
| compvc01 | vCenter default tenant | --                | NA                 | --                  | ⚠ Not Configured        | --             | compvc01 |
| mgmtvc01 | vCenter default tenant | --                | NA                 | --                  | ⚠ Not Configured        | --             | mgmtvc01 |

Figure 70 Tenants screen with vCenter tenants compvc01 and mgmtvc01

Segments are automatically created in BCF with VMware integration. To view the list of segments, select **Logical > Segments**.

| Name          | Tenant   | Description                        | Member VNI | Member VLAN | Interface Group Membership Rules | Switch Port Membership Rules | MAC Membership Rules | Endpoints | Segment Interface |
|---------------|----------|------------------------------------|------------|-------------|----------------------------------|------------------------------|----------------------|-----------|-------------------|
| compvc01-1621 | compvc01 | 1 vSphere portgroups: data2-comp   | --         | --          | 2                                | --                           | --                   | 2         | ⚠ Not Configured  |
| compvc01-1622 | compvc01 | 1 vSphere portgroups: vmotion-comp | --         | --          | 4                                | --                           | --                   | 4         | ⚠ Not Configured  |
| compvc01-1623 | compvc01 | 1 vSphere portgroups: vsan-comp    | --         | --          | 4                                | --                           | --                   | 4         | ⚠ Not Configured  |
| compvc01-1631 | compvc01 | 1 vSphere portgroups: data3-comp   | --         | --          | 1                                | --                           | --                   | 1         | ⚠ Not Configured  |
| mgmtvc01-1611 | mgmtvc01 | 1 vSphere portgroups: data1-mgmt   | --         | --          | 2                                | --                           | --                   | 2         | ⚠ Not Configured  |
| mgmtvc01-1612 | mgmtvc01 | 1 vSphere portgroups: vmotion-mgmt | --         | --          | 4                                | --                           | --                   | 4         | ⚠ Not Configured  |
| mgmtvc01-1613 | mgmtvc01 | 1 vSphere portgroups: vsan-mgmt    | --         | --          | 4                                | --                           | --                   | 4         | ⚠ Not Configured  |

Figure 71 Segments automatically created through vCenter integration

## 8.3 Configure segment interfaces

At this point, hosts and VMs on the same segments can communicate with each other. For communication between segments, segment interfaces are configured.

Tenants, segments, and segment interface addresses are shown in Table 15:

Table 15 BCF tenant and segment configuration

| Tenant   | Logical segment name | Function | VLAN ID | Subnet         | Segment interface address |
|----------|----------------------|----------|---------|----------------|---------------------------|
| mgmtvc01 | mgmtvc01-1611        | Data1    | 1611    | 172.16.11.0/24 | 172.16.11.254             |
|          | mgmtvc01-1612        | vMotion  | 1612    | 172.16.12.0/24 | 172.16.12.254             |
|          | mgmtvc01-1613        | vSAN     | 1613    | 172.16.13.0/24 | Not Used                  |
| compvc01 | compvc01-1621        | Data2    | 1621    | 172.16.21.0/24 | 172.16.21.254             |
|          | compvc01-1631        | Data3    | 1631    | 172.16.31.0/24 | 172.16.31.254             |
|          | compvc01-1622        | vMotion  | 1622    | 172.16.22.0/24 | 172.16.22.254             |
|          | compvc01-1623        | vSAN     | 1623    | 172.16.23.0/24 | Not Used                  |

To configure segment interfaces, do the following:

1. From the BCF GUI, go to **Logical > Tenants**.
2. Select a tenant, **mgmtvc01** in this example, to open the tenant configuration page.
3. In the left pane, scroll down and select **Segment Interfaces**. This adds **Segment Interfaces** to the right pane as shown in Figure 72.

**Note:** If additional items are selected in the left pane, you will need to scroll down the right pane to view the **Segment Interfaces** section.

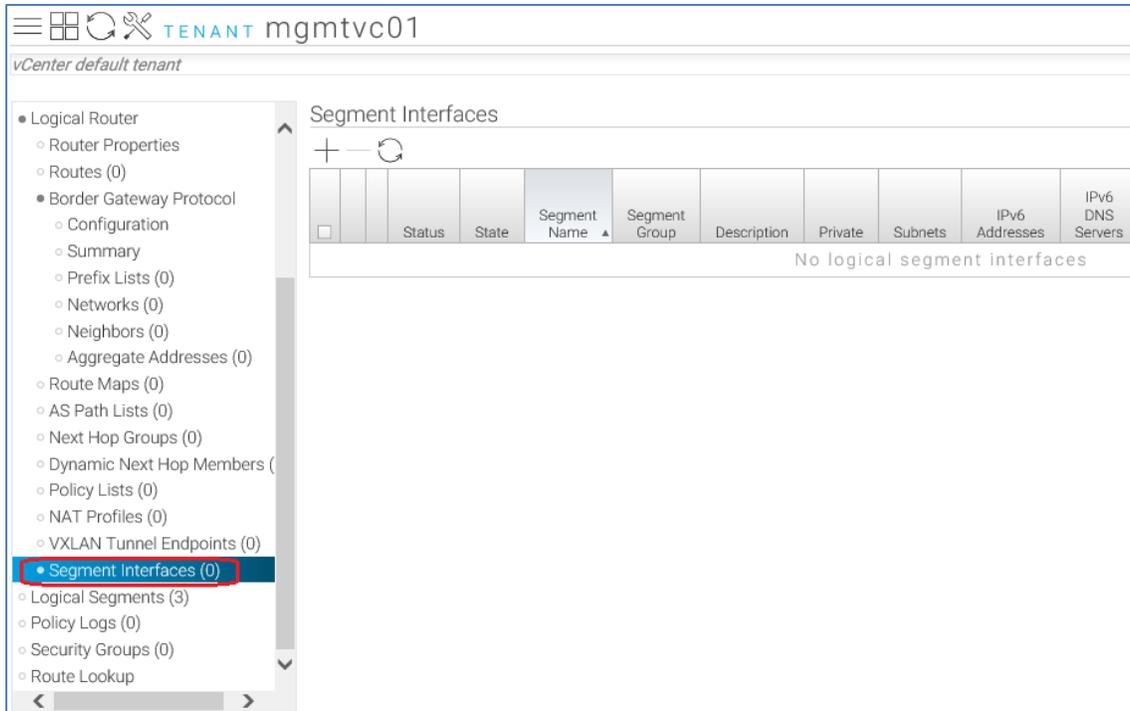


Figure 72 Segment Interfaces selected

- In the right pane under **Segment Interfaces**, click the **+** icon. The **Create Logical Segment Interface** dialog box displays.

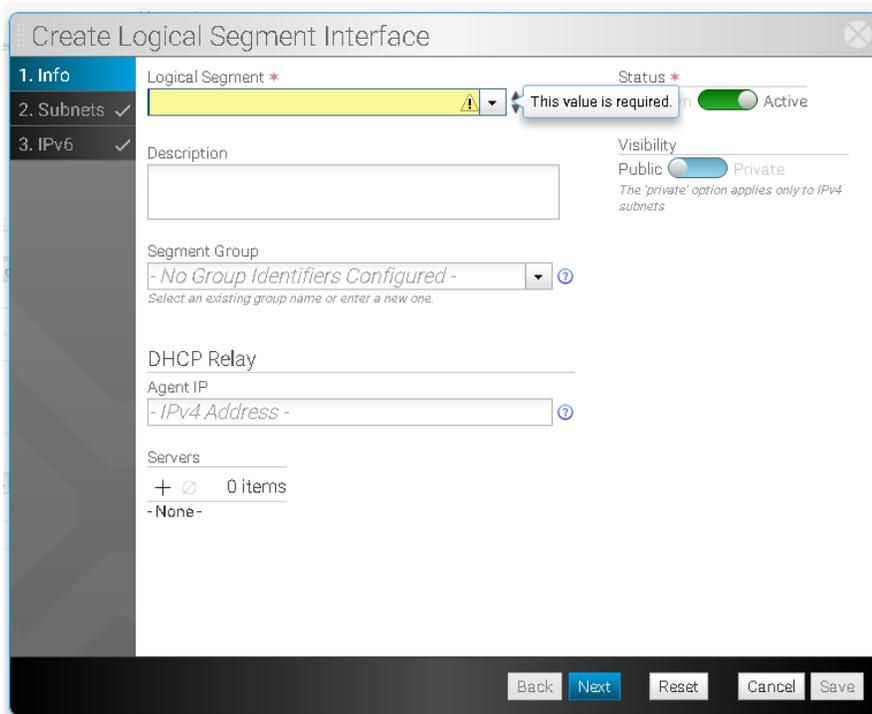


Figure 73 Create segment interface dialog box

5. Under **Logical Segment**, select the name of the first logical segment from the drop-down menu, **mgmtvc01-1611**. Leave other settings at their defaults and click **Next**.
6. Click the **+** icon to open the **Create Subnet** dialog box.
7. Provide the segment interface IP address and prefix per Table 15, **172.16.11.254 /24**. The subnet mask in dotted decimal form is automatically completed.

The screenshot shows a 'Create Subnet' dialog box. The 'IP CIDR \*' field is filled with '172.16.11.254' and a dropdown menu shows '24'. Below it, the subnet mask '255.255.255.0' is displayed. There is a 'Description' text box, a 'Virtual IP' section with a 'Virtual IP Address' dropdown showing '- IPv4 Address -' and a 'MAC Address' dropdown showing '- MAC Address -'. At the bottom right are 'Cancel' and 'Append' buttons.

Figure 74 Create subnet dialog box

8. Click **Append > Save**.

The first segment interface is created. Repeat the steps above to create segment interfaces for all remaining segments per Table 15.

**Note:** Since vSAN traffic is limited to a single segment, segment interfaces are not used for the two vSAN networks.

When complete, **Segment Interfaces** for mgmtvc01 and compvc01 appear as shown in Figure 75 and Figure 76.

| Segment Interfaces       |                          |                          |                          |        |        |                               |               |             |         |                  |                |
|--------------------------|--------------------------|--------------------------|--------------------------|--------|--------|-------------------------------|---------------|-------------|---------|------------------|----------------|
| + — ↻                    |                          |                          |                          |        |        |                               |               |             |         |                  |                |
| <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | Status | State  | Segment Name ▲                | Segment Group | Description | Private | Subnets          | IPv6 Addresses |
| <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | ✓ Up   | Active | <a href="#">mgmtvc01-1611</a> | —             | —           | —       | 172.16.11.254/24 | SLAAC          |
| <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | ✓ Up   | Active | <a href="#">mgmtvc01-1612</a> | —             | —           | —       | 172.16.12.254/24 | SLAAC          |

Figure 75 Mgmtvc01 segment interfaces configured

| Segment Interfaces       |                          |                          |                          |        |        |                               |               |             |         |                  |                |
|--------------------------|--------------------------|--------------------------|--------------------------|--------|--------|-------------------------------|---------------|-------------|---------|------------------|----------------|
| + — ↻                    |                          |                          |                          |        |        |                               |               |             |         |                  |                |
| <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | Status | State  | Segment Name ▲                | Segment Group | Description | Private | Subnets          | IPv6 Addresses |
| <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | ✓ Up   | Active | <a href="#">compvc01-1621</a> | —             | —           | —       | 172.16.21.254/24 | SLAAC          |
| <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | ✓ Up   | Active | <a href="#">compvc01-1622</a> | —             | —           | —       | 172.16.22.254/24 | SLAAC          |
| <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | ✓ Up   | Active | <a href="#">compvc01-1631</a> | —             | —           | —       | 172.16.31.254/24 | SLAAC          |

Figure 76 Compvc01 segment interfaces configured

## 8.4 Configure system tenant interfaces and logical router

At this point, VMs on different segments within the same tenant can communicate with each other, but not in different tenants.

For example, using Figure 69 as a reference, VM20 can ping VM30. However, VM20 cannot ping VM10 until the system tenant interfaces and logical router are configured.

To configure the System Tenant interfaces:

1. From the BCF GUI, go to **Logical > Tenants**.
2. Select a tenant, **mgmtvc01** in this example, to open the **Tenant** page.
3. In the left pane under **Logical Router**, click on **Router Properties** and **Routes** to display them in the right pane as shown in Figure 77:

The screenshot displays the BCF GUI for tenant **mgmtvc01**. The left navigation pane shows the **Logical Router** section expanded, with **Router Properties** and **Routes (8)** highlighted in red. The main right pane shows the **Logical Router Properties** for **mgmtvc01**. The **System Tenant Interface** is listed as **Not Configured**, with a red circle around the wrench icon. Below this, the **Routes** table is visible, showing 8 routes with their respective configurations.

|                          | Configured | Preference | Description | CIDR            | Type      | Next Hop Tenant |
|--------------------------|------------|------------|-------------|-----------------|-----------|-----------------|
| <input type="checkbox"/> | -          | 0          | -           | 172.16.12.2/32  | Host      | mgmtvc01        |
| <input type="checkbox"/> | -          | 0          | -           | 172.16.12.1/32  | Host      | mgmtvc01        |
| <input type="checkbox"/> | -          | 0          | -           | 172.16.11.11/32 | Host      | mgmtvc01        |
| <input type="checkbox"/> | -          | 0          | -           | 172.16.11.10/32 | Host      | mgmtvc01        |
| <input type="checkbox"/> | -          | 0          | -           | 172.16.12.4/32  | Host      | mgmtvc01        |
| <input type="checkbox"/> | -          | 0          | -           | 172.16.12.3/32  | Host      | mgmtvc01        |
| <input type="checkbox"/> | -          | 0          | -           | 172.16.11.0/24  | Connected | mgmtvc01        |
| <input type="checkbox"/> | -          | 0          | -           | 172.16.12.0/24  | Connected | mgmtvc01        |

Figure 77 Logical Router properties and Routes

4. In the right pane under **Logical Router Properties**, click the  icon next to **System Tenant Interface** (circled in red in Figure 77). The **Manage Tenant Interfaces** dialog box displays.
5. Move the **Configured** and **Active** sliders to the right to enable the interfaces.

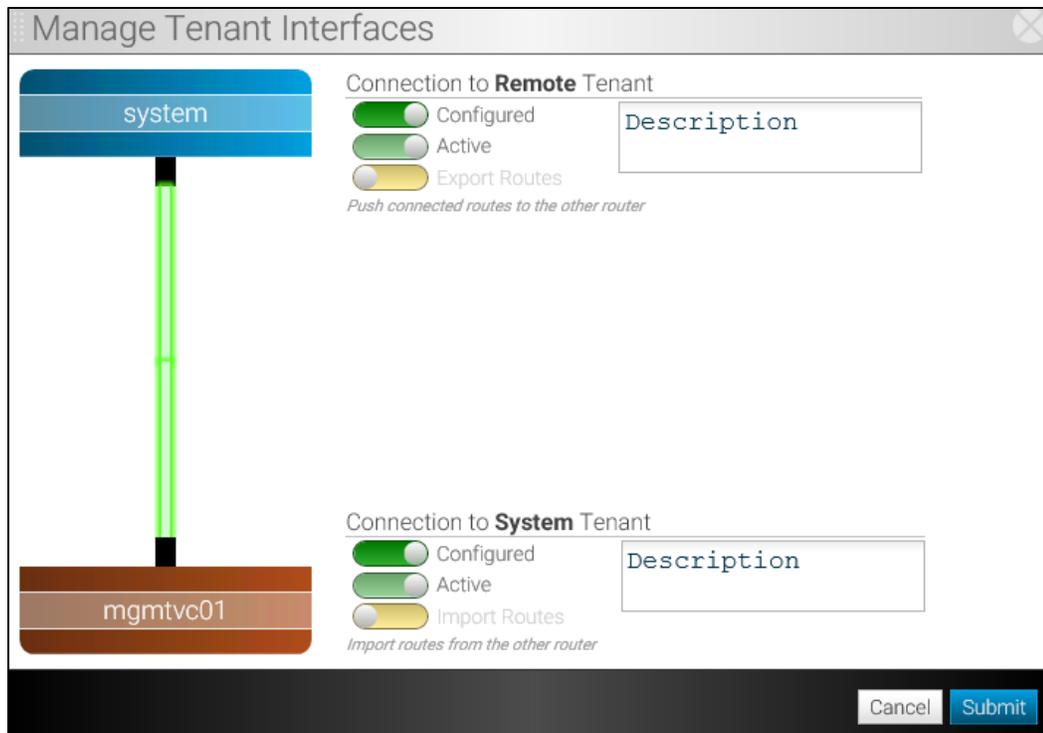


Figure 78 Enabling system tenant interfaces

6. Click **Submit** to apply the changes. The **System Tenant Interface** displays as **Up** under **Logical Router Properties**.
7. Scroll down to the **Routes** section and click the **+** icon to open the **Create Route** dialog box.

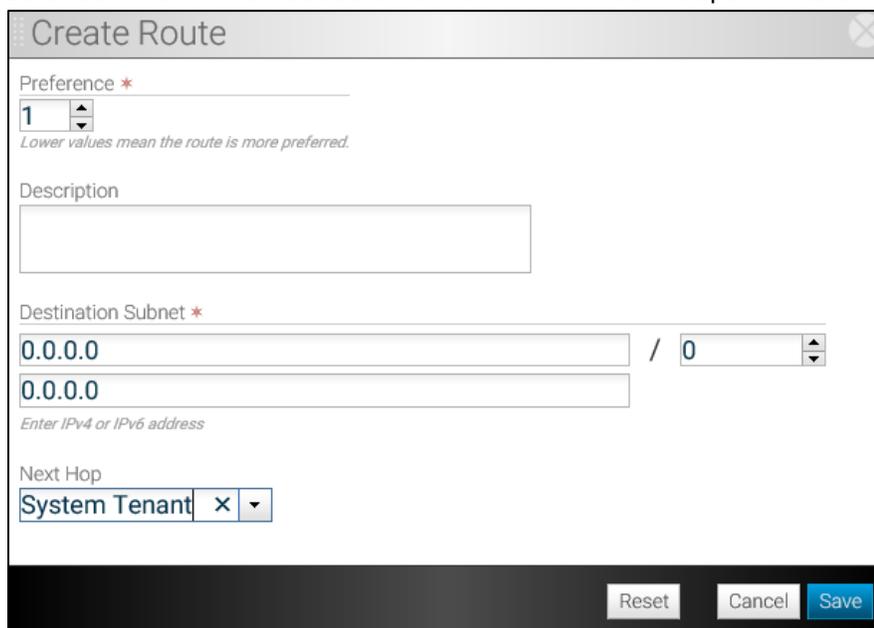


Figure 79 Create route dialog box

- In this example, the destination subnet is set to **0.0.0.0/0** (any). Set **Next Hop** to **System Tenant** and click **Save**.

For the **mgmtvc01** tenant, **Logical Router Properties** and **Routes** appears as shown in Figure 80.

The screenshot shows the configuration for the **mgmtvc01** logical router. The **Logical Router Properties** section includes:

- MAC Address: 5c:16:c7:0a:33:1d
- Applied Policy List: - None -
- VRF ID: 1
- System Tenant Interface: ✓ Up
- Default Route: system tenant
- vCenter: mgmtvc01

The **Routes** table lists the following entries:

| Configured                          | Preference | Description | CIDR            | Type      | Next Hop Tenant | Next Hop Group              | Next Hop IP Address | Status |
|-------------------------------------|------------|-------------|-----------------|-----------|-----------------|-----------------------------|---------------------|--------|
| <input type="checkbox"/>            | 0          |             | 172.16.12.2/32  | Host      | mgmtvc01        | Segment Iface mgmtvc01-1612 | -                   | Active |
| <input type="checkbox"/>            | 0          |             | 172.16.12.1/32  | Host      | mgmtvc01        | Segment Iface mgmtvc01-1612 | -                   | Active |
| <input type="checkbox"/>            | 0          |             | 172.16.11.11/32 | Host      | mgmtvc01        | Segment Iface mgmtvc01-1611 | -                   | Active |
| <input type="checkbox"/>            | 0          |             | 172.16.11.10/32 | Host      | mgmtvc01        | Segment Iface mgmtvc01-1611 | -                   | Active |
| <input type="checkbox"/>            | 0          |             | 172.16.12.4/32  | Host      | mgmtvc01        | Segment Iface mgmtvc01-1612 | -                   | Active |
| <input type="checkbox"/>            | 0          |             | 172.16.12.3/32  | Host      | mgmtvc01        | Segment Iface mgmtvc01-1612 | -                   | Active |
| <input type="checkbox"/>            | 0          |             | 172.16.11.0/24  | Connected | mgmtvc01        | Segment Iface mgmtvc01-1611 | -                   | Active |
| <input type="checkbox"/>            | 0          |             | 172.16.12.0/24  | Connected | mgmtvc01        | Segment Iface mgmtvc01-1612 | -                   | Active |
| <input checked="" type="checkbox"/> | 1          |             | ::/0            | Static    | system          | Tenant iface system         | -                   | Active |
| <input checked="" type="checkbox"/> | 1          |             | 0.0.0.0/0       | Static    | system          | Tenant iface system         | -                   | Active |

Figure 80 mgmtvc01 tenant logical router properties and routes

Repeat the steps above for the **compvc01** tenant. When done its page appears as shown in Figure 81.

The screenshot shows the configuration for the **compvc01** logical router. The **Logical Router Properties** section includes:

- MAC Address: 5c:16:c7:0a:33:1d
- Applied Policy List: - None -
- VRF ID: 2
- System Tenant Interface: ✓ Up
- Default Route: system tenant
- vCenter: compvc01

The **Routes** table lists the following entries:

| Configured                          | Preference | Description | CIDR            | Type      | Next Hop Tenant | Next Hop Group              | Next Hop IP Address | Status |
|-------------------------------------|------------|-------------|-----------------|-----------|-----------------|-----------------------------|---------------------|--------|
| <input type="checkbox"/>            | 0          |             | 172.16.31.30/32 | Host      | compvc01        | Segment Iface compvc01-1631 | -                   | Active |
| <input type="checkbox"/>            | 0          |             | 172.16.22.1/32  | Host      | compvc01        | Segment Iface compvc01-1622 | -                   | Active |
| <input type="checkbox"/>            | 0          |             | 172.16.21.21/32 | Host      | compvc01        | Segment Iface compvc01-1621 | -                   | Active |
| <input type="checkbox"/>            | 0          |             | 172.16.22.2/32  | Host      | compvc01        | Segment Iface compvc01-1622 | -                   | Active |
| <input type="checkbox"/>            | 0          |             | 172.16.22.3/32  | Host      | compvc01        | Segment Iface compvc01-1622 | -                   | Active |
| <input type="checkbox"/>            | 0          |             | 172.16.21.20/32 | Host      | compvc01        | Segment Iface compvc01-1621 | -                   | Active |
| <input type="checkbox"/>            | 0          |             | 172.16.22.4/32  | Host      | compvc01        | Segment Iface compvc01-1622 | -                   | Active |
| <input type="checkbox"/>            | 0          |             | 172.16.22.0/24  | Connected | compvc01        | Segment Iface compvc01-1622 | -                   | Active |
| <input type="checkbox"/>            | 0          |             | 172.16.21.0/24  | Connected | compvc01        | Segment Iface compvc01-1621 | -                   | Active |
| <input type="checkbox"/>            | 0          |             | 172.16.31.0/24  | Connected | compvc01        | Segment Iface compvc01-1631 | -                   | Active |
| <input checked="" type="checkbox"/> | 1          |             | 0.0.0.0/0       | Static    | system          | Tenant iface system         | -                   | Active |
| <input checked="" type="checkbox"/> | 1          |             | ::/0            | Static    | system          | Tenant iface system         | -                   | Active |

Figure 81 compvc01 tenant logical router properties and routes

## 8.5 Verifying connectivity

This section shows the ping syntax used to validate connectivity. The vSAN and vMotion tests are done from the ESXi CLI. The VM test is done within the guest OS.

### 8.5.1 vSAN networks

On the vSAN networks in this deployment, ESXi hosts must be able to reach other ESXi hosts in the same cluster. Hosts should not be able to reach hosts in other clusters. The vSAN network should be able to support 9000 byte (jumbo) frames.

In the following example, host comp01 successfully pings host comp02's vSAN VMkernel IP address:

```
[root@comp01:~] ping 172.16.23.2 -d -s 8950
PING 172.16.23.2 (172.16.23.2): 8950 data bytes
8958 bytes from 172.16.23.2: icmp_seq=0 ttl=64 time=0.495 ms
8958 bytes from 172.16.23.2: icmp_seq=1 ttl=64 time=0.537 ms
8958 bytes from 172.16.23.2: icmp_seq=2 ttl=64 time=0.507 ms
```

The `-d` argument means don't fragment the packet and the `-s 8950` argument sets the ICMP data size in the packet. (This size does not include IP headers, so it is set to slightly under 9000).

**Note:** If pings fail with jumbo frames, check the MTU size settings on the associated VDS and VMkernel adapters in vCenter.

### 8.5.2 vMotion networks

On the vMotion networks in this deployment, ESXi hosts must be able to reach all other ESXi hosts in the Management and Compute clusters. The vMotion network should be able to support 9000 byte frames.

In the following example, host comp01 successfully pings the vMotion VMkernel IP address of host mgmt01:

```
[root@comp01:~] ping 172.16.12.1 -S vmotion -s 8950 -d
PING 172.16.12.1 (172.16.12.1): 8950 data bytes
8958 bytes from 172.16.12.1: icmp_seq=0 ttl=61 time=0.445 ms
8958 bytes from 172.16.12.1: icmp_seq=1 ttl=61 time=0.469 ms
8958 bytes from 172.16.12.1: icmp_seq=2 ttl=61 time=0.426 ms
```

The `-S vmotion` argument instructs the utility to use the vMotion TCP/IP stack. This argument is required for the command to succeed.

The `-d` argument means don't fragment the packet and the `-s 8950` argument sets the ICMP data size in the packet. (This size does not include IP headers, so it is set to slightly under 9000).

**Note:** If pings fail with jumbo frames, check the MTU size settings on the associated VDS and VMkernel adapters in vCenter.

### 8.5.3 VM networks

To test VM communication, the vSphere Web Client is used to add a vNIC to each VM on its assigned data network. For example, VM10 in the Management cluster has a network adapter configured on the Data1 network. Its vNIC is configured as shown in Figure 82.

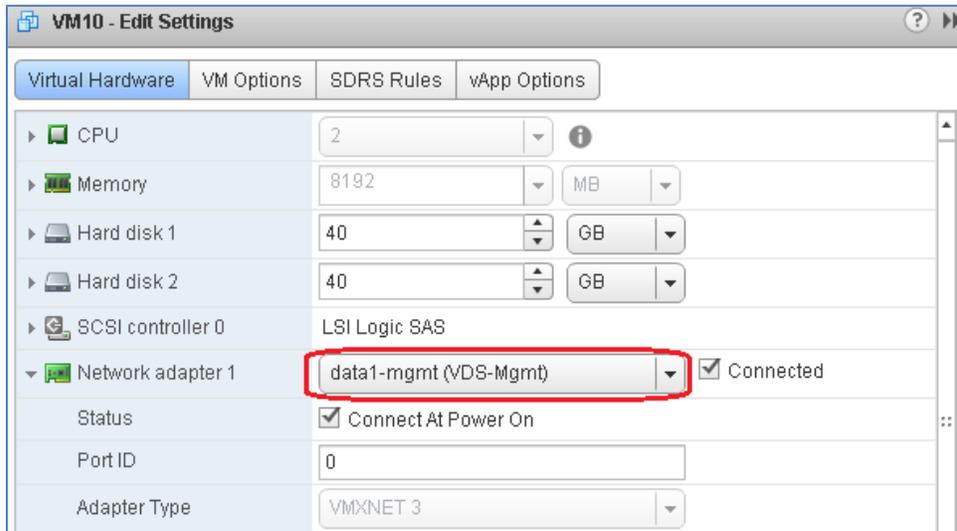


Figure 82 vNIC settings on VM10

In the VM10 guest OS, the IP address and gateway are configured on the Data1 network using the IP addressing shown in Table 15. The other VMs are configured in the same manner on their respective data networks.

With guest OS firewalls disabled or configured to allow ping responses, VM10 is able to successfully ping all other VMs on the Data1, Data2, and Data3 networks. In Figure 83, VM10 in the Management cluster successfully pings VM30 in the Compute cluster.

```
C:\Windows\system32>ping 172.16.31.30

Pinging 172.16.31.30 with 32 bytes of data:
Reply from 172.16.31.30: bytes=32 time<1ms TTL=125

Ping statistics for 172.16.31.30:
    Packets: Sent = 4, Received = 4, Lost = 0 (0% loss),
    Approximate round trip times in milli-seconds:
        Minimum = 0ms, Maximum = 0ms, Average = 0ms

C:\Windows\system32>
```

Figure 83 VM10 pings VM30

## 9 Create vSAN clusters

This section provides a brief outline of the steps to prepare for and create vSAN clusters for this deployment. For vSAN resources, see Appendix D.

Servers used in vSANs must either have a mix of flash (SSD) and magnetic (HDD) drives or be all-flash. See Appendix B for the servers and disks used in this deployment and the [VMware vSAN Design and Sizing Guide](#) for storage requirements and guidance.

For redundancy, vSANs employ software RAID. With the exception of single drive RAID-0 configurations, vSANs do not support hardware RAID. The PowerEdge R740xd servers used in this deployment each have HBA330 controllers which do not support hardware RAID. The PowerEdge R630 servers have PERC H730 cards which support hardware RAID, but are set to HBA (non-RAID) mode. See your system documentation to ensure storage controllers are not set to RAID mode for disks that will be part of a vSAN.

vSAN is enabled on each cluster by following the instructions in the [VMware vSAN Operations Guide](#). When creating the Compute cluster vSAN, the network validation page appears as shown in Figure 84. This confirms that VMware and BCF are properly configured for vSAN network functionality in the Compute cluster. The network validation page for the Management Cluster is similar.

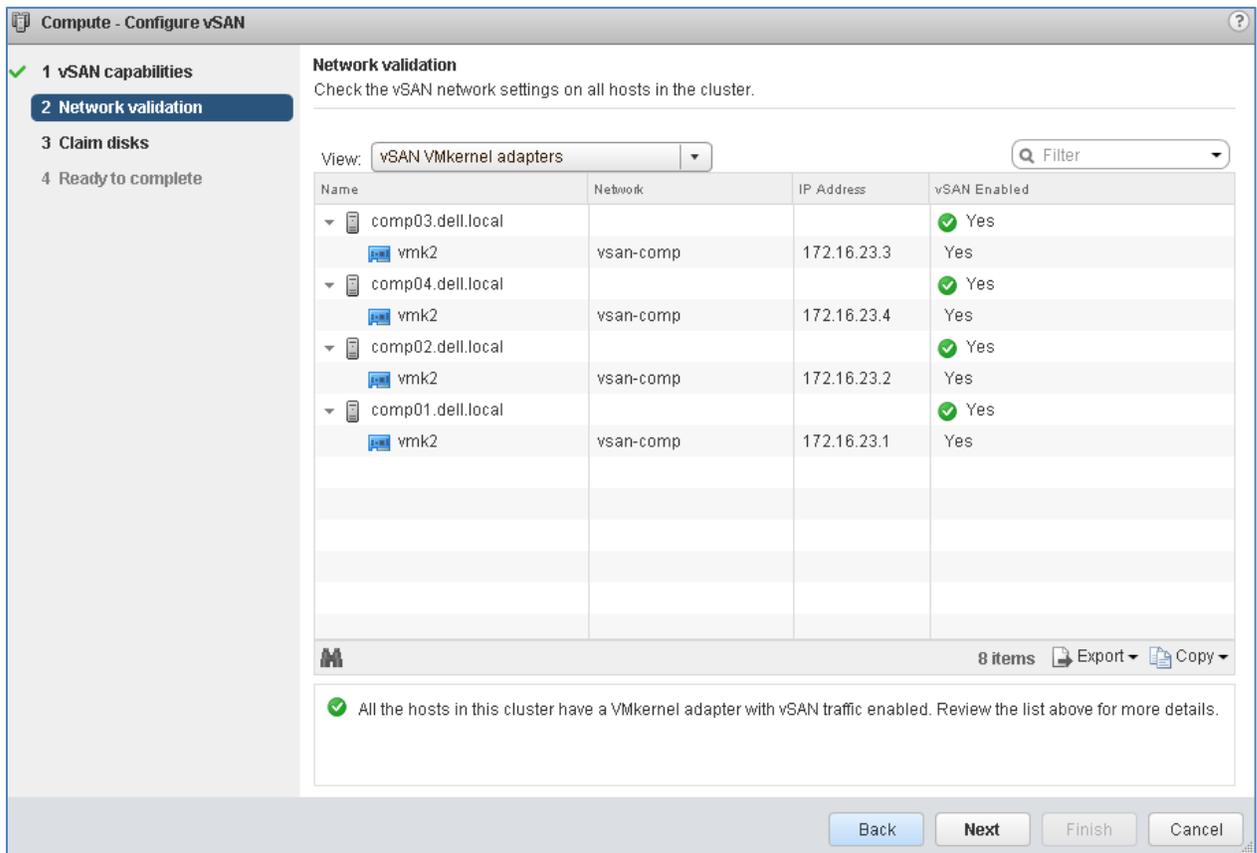


Figure 84 vSAN network validation page – Compute cluster

After vSANs are created in each cluster, a datastore named **vsanDatastore** is listed under each data center on the **Navigation pane > Storage tab** similar to that shown in Figure 85.

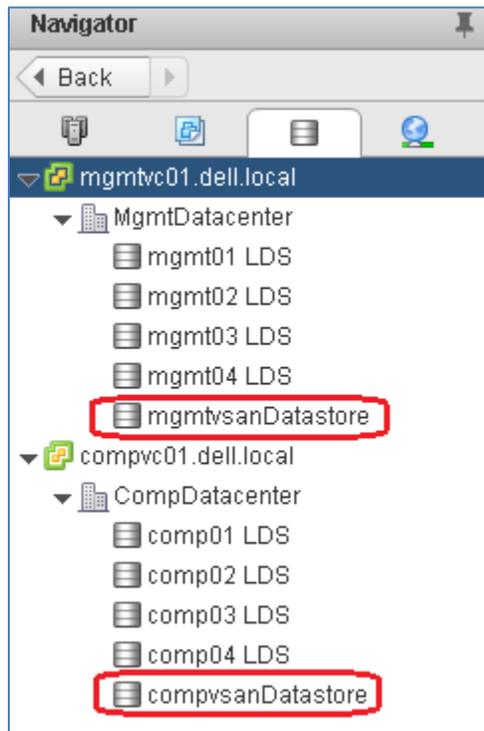


Figure 85 vSAN datastores created

The vSAN datastores in Figure 85 have been renamed to **mgmtvsanDatastore** and **compvsanDatastore** for usability. Other datastores shown in Figure 85 are local datastores (named *hostname##* LDS) on each host. Until vSANs are created, all VMs initially reside on local datastores.

The next steps are to:

1. Enable vSphere DRS and HA on each cluster.
2. Migrate VMs from local datastores to vSAN datastores using vMotion to take advantage of DRS and HA features.

**Note:** Refer to [VMware Documentation](#) online for more information on DRS, HA, and VM migration.

## A Rack diagrams

The racks and equipment used in this deployment guide are shown in Figure 86.

Each rack contains one S3048-ON OOB management switch and two S4048-ON leaf switches. Rack 1 also contains the Z9100-ON spine switches, the Management cluster (four R630 hosts), and two BCF Controllers. The Compute cluster in Rack 2 contains four R740xd hosts with capacity for fifteen additional R740xd hosts.

Compute racks are added as needed with up to nineteen R740xd hosts per rack. An additional compute cluster is created in vCenter for each additional rack.

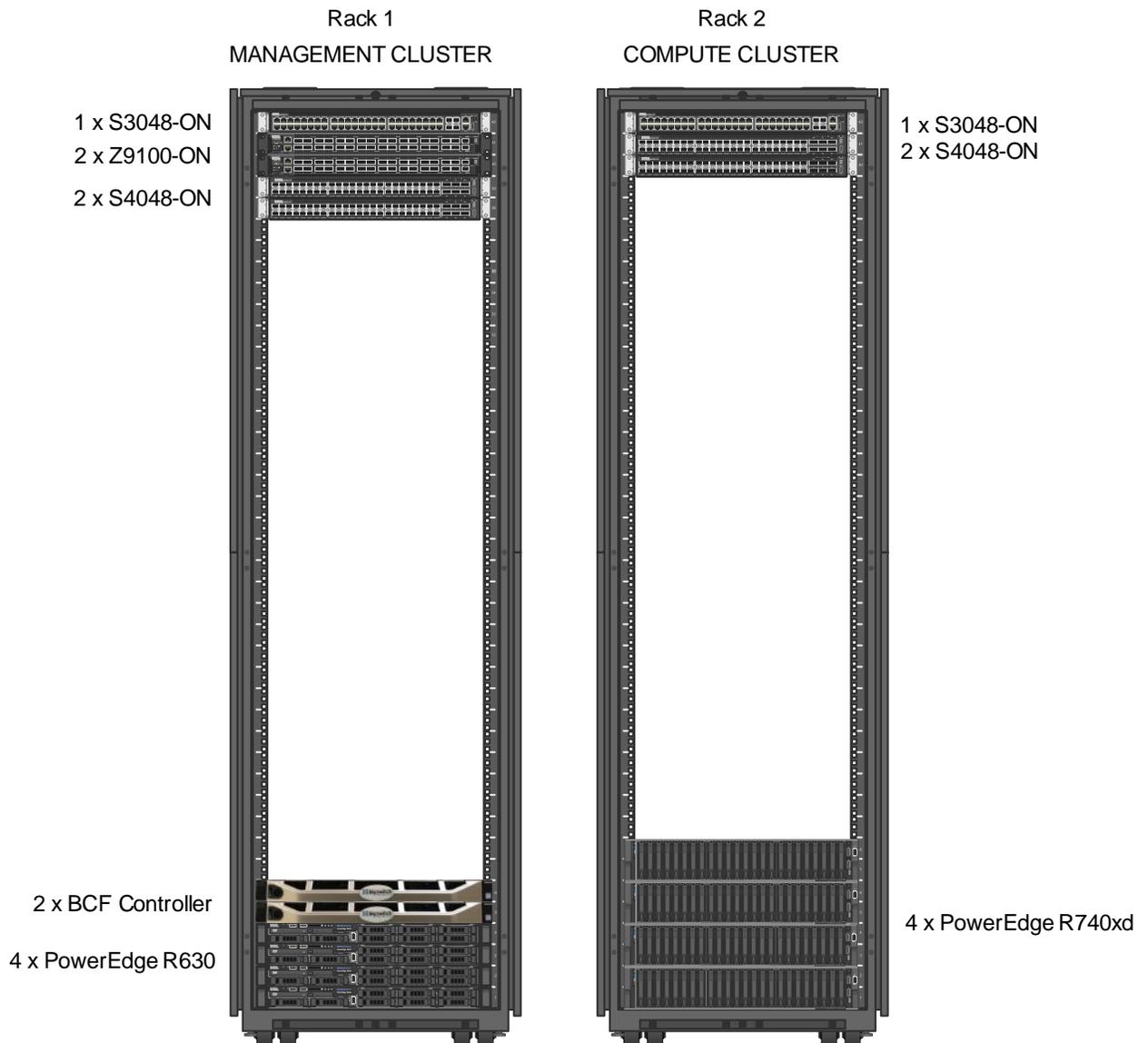


Figure 86 Racks and equipment used in this guide

## B Dell EMC validated hardware and components

The following tables list the hardware and components used to configure and validate the example configurations in this guide.

### B.1 Switches

| Qty | Item   | OS/Firmware version           |
|-----|--|-------------------------------|
| 2   | S3048-ON Management switch   | <b>OS:</b> DNOS 9.11.2.5      |
|     |  | <b>System CPLD:</b> 9         |
|     |  | <b>Module CPLD:</b> 7         |
| 4   | S4048-ON Leaf switch<br><br>Switch Light OS and CPLD/ONIE firmware are provided by controller running BCF 4.2.3  | <b>OS:</b> Switch Light 4.2.3 |
|     |  | <b>CPLD:</b> 15.12.5          |
|     |  | <b>ONIE:</b> 3.21.1.2         |
| 2   | Z9100-ON Spine switch<br><br>Switch Light OS and CPLD/ONIE firmware are provided by controller running BCF 4.2.3 | <b>OS:</b> Switch Light 4.2.3 |
|     |  | <b>CPLD:</b> 6.4.4.4          |
|     |  | <b>ONIE:</b> 3.23.1.3         |

### B.2 PowerEdge R740xd servers

This guide uses four PowerEdge R740xd servers in the Compute cluster.

| Qty per server | Item  | Firmware version                   |
|----------------|---|------------------------------------|
| 2              | Intel Xeon Gold 6130 CPU @ 2.10GHz, 16 cores                | -                                  |
| 64             | GB RAM  | -                                  |
| 20             | 400GB SAS SSD   | -                                  |
| 1              | Dell HBA330 Storage Controller                              | 13.17.03.00                        |
| 1              | Broadcom QP rNDC: 5720 DP 1GbE Base-T + 57412 DP 10GbE SFP+ | 5720: 7.10.0<br>57412: 20.06.05.06 |
| 1              | Mellanox ConnectX-4 LX 25GbE (with 10Gb SFP+ modules)       | 14.17.20.52                        |
| -              | BIOS  | 1.0.7                              |
| -              | iDRAC with Lifecycle Controller                             | 3.00.00.00                         |

## B.3 PowerEdge R630 servers

This guide uses four PowerEdge R630 servers in the Management cluster.

| Qty per server | Item                                       | Firmware Version |
|----------------|--|------------------|
| 1              | Intel Xeon E5-2650 v3 2.3GHz CPU, 10 cores | -                |
| 128            | GB RAM                                     | -                |
| 2              | 400GB SATA SSD                             | -                |
| 6              | 300GB SAS HDD                              | -                |
| 1              | PERC H730 Mini Storage Controller          | 25.5.3.0005      |
| 1              | QLogic 57840 10GbE QP rNDC                 | 10.01.00         |
| 1              | Intel I350-T Base-T 1GbE DP PCIe adapter   | 18.0.17          |
| -              | BIOS                                       | 2.4.3            |
| -              | iDRAC with Lifecycle Controller            | 2.30.30.30       |

## C Dell EMC validated software and required licenses

The Software table lists the software components used to validate the example configurations in this guide. The VMware Licenses section lists the VMware licenses required for the example configurations in this guide.

### C.1 Software

| Item                            | Version  |
|---------------------------------|--|
| Big Cloud Fabric                | 4.2.3  |
| VMware ESXi                     | 6.5 U1 - Dell EMC customized image version A00 build 5969303 |
| VMware vCenter Server Appliance | 6.5 U1 – build 5973321                                       |
| VMware vSAN                     | 6.6.1 (provided with ESXi 6.5 U1 build 5969303)              |

### C.2 VMware Licenses

vCenter Servers are licensed by instance. The remaining licenses are allocated based on the number of CPU sockets in the participating hosts.

Required licenses for the topology built in this guide are as follows:

- VMware vSphere 6 Enterprise Plus – 20 CPU sockets
- VMware vCenter 6 Server Standard – 2 instances
- VMware vSAN Standard – 16 CPU sockets

VMware product licenses are centrally managed by going to the **vSphere Web Client Home** page and selecting **Licensing** in the center pane.

## D Technical support and resources

### D.1 Dell EMC product manuals and technical guides

[Dell EMC TechCenter](#) - An online technical community where IT professionals have access to numerous resources for Dell EMC software, hardware and services.

[Dell EMC TechCenter Networking Guides](#)

[Dell EMC Ready Bundle for Virtualization Datasheet](#)

[Manuals and documentation for Dell EMC Networking S3048-ON](#)

[Manuals and documentation for Dell EMC Networking S4048-ON](#)

[Manuals and documentation for Dell EMC Networking Z9100-ON](#)

### D.2 Big Switch Networks product manuals and technical guides

[Big Switch Networks support site](#) – Contains support information and guides referenced in this paper.

**Note:** An account is required to use the support site. Contact your Big Switch Networks account representative.

[Big Cloud Fabric Datasheet](#)

[Big Cloud Fabric: A Next-Generation Data Center Switching Platform](#)

[Big Switch Networks + Dell: Ideal SDN Fabric for VMware SDDC](#)

### D.3 VMware product manuals and technical guides

[VMware Documentation](#)

[VMware Compatibility Guide](#)

[VMware vSAN 6.6 Datasheet](#)

[VMware vSAN Technical Resources](#)

[VMware vSAN Design and Sizing Guide](#)

[VMware vSAN Operations Guide](#)

[VMware Validated Designs Documentation](#) – Release 4.1

[vSphere Installation and Setup](#) – This document includes ESXi 6.5 and vCenter Server 6.5.

## E Support and feedback

### Contacting Technical Support

Support Contact Information

Web: <http://support.dell.com/>

Telephone: USA: 1-800-945-3355

### Feedback for this document

We encourage readers to provide feedback on the quality and usefulness of this publication by sending an email to [Dell\\_Networking\\_Solutions@Dell.com](mailto:Dell_Networking_Solutions@Dell.com).