Cisco Dynamic Fabric Automation Technology: Overview and Deployment Considerations

What You Will Learn

This document introduces the Cisco Nexus® unified fabric technology innovation called Cisco® Dynamic Fabric Automation (DFA). Cisco DFA helps simplify, optimize, and automate the unified fabric environment by offering an architecture based on four main building blocks. Each of these building blocks provides a group of functions that can be deployed independently in a modular fashion. This approach allows easy adoption of new technology as the data center fabric architecture evolves.

As of this writing, the Cisco Nexus 7000 and 6000 Series Switches, Cisco Nexus 5500 as well as 5600 series, Cisco Nexus 1000V Switch, and Cisco Prime™ Data Center Network Manager (DCNM) are the Cisco products that can be used in a unified fabric running Cisco DFA. This document presents technology and deployment considerations for simplifying, optimizing, and automating a data center fabric built with Cisco Nexus switches and Cisco Prime DCNM.

For more information about hardware, software versions, and feature support, see Table 1.

Table 1. Cisco DFA Hardware and Software Requirements

<table>
<thead>
<tr>
<th>Function</th>
<th>Platform</th>
<th>I/O Module</th>
<th>Minimum Software Version</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cisco DFA spine</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cisco Nexus 7000 Series</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cisco Nexus 7000 Series</td>
<td>Cisco Nexus F3 Module</td>
<td>Release 6.2(6a)</td>
<td></td>
</tr>
<tr>
<td>Cisco Nexus 7000 Series</td>
<td>Cisco Nexus F2 and F2e Module</td>
<td>Release 6.2(6a)</td>
<td></td>
</tr>
<tr>
<td>Cisco Nexus 7000 Series</td>
<td>Cisco Nexus F3 Module</td>
<td>Release 6.2(6a)</td>
<td></td>
</tr>
<tr>
<td>Cisco Nexus 6000 Series</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cisco Nexus 6000 Series</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cisco Nexus 6000 Series</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cisco Nexus 6000 Series</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cisco Nexus 6000 Series</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cisco Nexus 5600 Series</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cisco Nexus 5600 Series</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cisco Nexus 5600 Series</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cisco Nexus 5600 Series</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cisco DFA leaf</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cisco Nexus 7000 Series</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cisco Nexus 7000 Series</td>
<td>Cisco Nexus F3 Module</td>
<td>Release 7.1(x)</td>
<td></td>
</tr>
<tr>
<td>Cisco Nexus 7000 Series</td>
<td>Cisco Nexus F3 Module</td>
<td>Release 7.1(x)</td>
<td></td>
</tr>
<tr>
<td>Cisco Nexus 7000 Series</td>
<td>Cisco Nexus F3 Module</td>
<td>Release 7.1(x)</td>
<td></td>
</tr>
<tr>
<td>Cisco Nexus 7000 Series</td>
<td>Cisco Nexus F3 Module</td>
<td>Release 7.1(x)</td>
<td></td>
</tr>
<tr>
<td>Cisco Nexus 6000 Series</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cisco Nexus 6000 Series</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cisco Nexus 6000 Series</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cisco Nexus 6000 Series</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cisco Nexus 5600 Series</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cisco Nexus 5600 Series</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cisco DFA border leaf</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cisco Nexus 7000 Series</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cisco Nexus 7000 Series</td>
<td>Cisco Nexus F3 Module</td>
<td>Release 7.1(x)</td>
<td></td>
</tr>
<tr>
<td>Cisco Nexus 7000 Series</td>
<td>Cisco Nexus F3 Module</td>
<td>Release 7.1(x)</td>
<td></td>
</tr>
<tr>
<td>Cisco Nexus 7000 Series</td>
<td>Cisco Nexus F3 Module</td>
<td>Release 7.1(x)</td>
<td></td>
</tr>
<tr>
<td>Cisco Nexus 6000 Series</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cisco Nexus 6000 Series</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cisco Nexus 6000 Series</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cisco Nexus 6000 Series</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cisco Nexus 5600 Series</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cisco Nexus 5600 Series</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cisco DFA Layer 2 leaf</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cisco Nexus 5500 Series</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cisco Nexus 5500 Series</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Fabric management</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cisco DCNM</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cisco DCNM</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Virtual Station Interface (VSI)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Discovery and Configuration Protocol (VDP) signaling</td>
<td>Cisco Nexus 1000V</td>
<td>VMware vSphere 5.1 and 5.5</td>
<td>Release 4.2(1)SV2(2.2)</td>
</tr>
</tbody>
</table>

Note: Cisco DFA requires at least one Multiprotocol Border Gateway Protocol (MP-BGP) route reflector. This is an integrated function of Cisco DFA, and the Cisco Nexus 6000 Series as well as the Nexus 5600 Series supports this function using the minimum software release listed in Table 1. The Cisco Nexus 7000 Series and as well the Cisco Nexus 7700 Series supports the Cisco DFA MP-BGP route reflector as of Release 6.2(6a).
As necessary, this document identifies Cisco DFA features for specific switching platforms as available in the current release or in a planned future release.

Cisco DFA Building Blocks

Cisco DFA consists of four building blocks (Figure 1):

- Fabric management
- Workload automation
- Optimized networking
- Virtual fabrics

Figure 1. Dynamic Fabric Automation Architecture: Innovative Building Blocks

Fabric Management

Fabric management is a fundamental building block of the Cisco DFA fabric provided through a collection of capabilities combined in the Cisco Prime DataCenter Network Manager or DCNM. These capabilities provide new levels of management, simplifying workload visibility, optimizing troubleshooting, and automating fabric component provisioning.

The Cisco DFA simplified Fabric Management is achieved by the following Cisco products and well known network services:

- Cisco Prime DCNM
- Dynamic Host Configuration Protocol (DHCP)
- Trivial File Transfer Protocol (TFTP)
- Secure Copy Protocol (SCP)
- Extensive Message Protocol (XMPP)
- Lightweight Directory Access Protocol (LDAP)
- Advanced Message Queuing Protocol (AMQP) message broker
- Cisco Prime Network Services Controller (PNSC) Adaptor

On top of the fabric access layer, the physical switches are initially provisioned using Power-On Auto Provisioning (POAP) with fully integrated, easy-to-use templates and workflow. As the fabric is deployed, the cable-plan consistency check builds a level of verification to check the network configuration and the related physical cabling.

Fabric management also provides extensive visibility into the physical network: from the virtual fabric to the physical or virtual server.
**Workload Automation**

Workload automation integrates with automation and orchestration tools through northbound REST APIs in JSON format. Cisco DFA provides a predefined and open programmatic interface for OpenStack and Cisco UCS® Director by using REST APIs. VMWare vCloud Director uses the AMQP over a packaged script to interconnect with Cisco Prime Data Center Network Manager (DCNM) in a different decoupled and open way.

For additional information, please see the “Simplified Fabric Management” white paper or the API and programmatic reference documents, which can be found at [http://www.cisco.com/go/dfa](http://www.cisco.com/go/dfa).

In addition, Cisco DFA offers a level of control for fabric component provisioning through the automatic application of templates (POAP) and profiles (network autoconfiguration), using southbound APIs and standards-based protocols. This mechanism can also be extended to Layer 4 to 7 network services such as firewalls and load balancers.

Workload automation provides the capability to guide configuration of fabric components based on the requirements and location of a physical or virtual server. This capability enables the network to address the operation demands of workload mobility and provide any workload, anywhere, dynamically. Northbound APIs provide an interface for orchestration tools for consistent workload, network, and service provisioning based on network and network-service profiles and policies. Network and network-service autoconfiguration together with the open APIs and connected orchestrator provides automated provisioning whenever a given workload requires it. Workload automation is built on top of the Cisco Prime DCNM component provided in Cisco DFA.

**Optimized Networking**

Building a data center fabric has traditionally involved a trade-off between the flexibility of forwarding Ethernet frames at Layer 2 (switching), and the stability and small failure domains of forwarding IP packets at Layer 3 (routing). Optimized networking allows the Cisco DFA fabric to offer the best attributes of Layer 2 switching and Layer 3 routing concurrently; the hard trade-off decisions no longer need to be made. Cisco DFA optimized networking results in small failure domains, with any IP subnet supported anywhere on the fabric concurrently through the use of a simple distributed default gateway mechanism. Redundant switching models for spine and leaf elements also provide N+ redundancy across the entire fabric. Other properties of optimized networking do include high bisectional bandwidth as well as uniform reachability and deterministic latency.

Optimized networking uses Cisco FabricPath frame encapsulation for efficient forwarding based on a Shortest Path First (SPF) algorithm for unicast and multicast IP traffic. Host and subnet route distribution across the fabric is accomplished using a scalable MP-BGP control plane.

Cisco DFA enhances the traditional frame and packet forwarding by improving the Ethernet concept of flood and learn. By using Cisco FabricPath frame encapsulation and its conversational learning for Layer 2 traffic, Cisco DFA uses the forwarding methodology of frame routing to overcome logical singletree topologies. By adding a control plane for host and subnet reachability, unnecessary flooding can be reduced by using pro-active learning.

In addition to the enhanced control plane and data plane for unicast and multicast forwarding, Cisco DFA reduces the Layer 2 failure domain by having the Layer 2 and Layer 3 demarcation on the host-connected leaf switch, terminating the host-originated discovery protocols at this layer.

The use of the distributed proxy or anycast gateway technology on all leaf switches for a given VLAN improves resilience and allows the fabric to scale to more hosts by keeping a short path for forwarding within and between VLANs.
A major difference between the distributed gateway technology in Cisco DFA and anycast Hot Standby Router Protocol (HSRP) or traditional other First-Hop Redundancy Protocols (FHRPs) such as HSRP, Virtual Router Redundancy Protocol (VRRP), and Gateway Load Balancing Protocol (GLBP) is the absence of a hello exchange between the various leaves participating and serving the same virtual IP address. Each Cisco DFA leaf provides the same gateway MAC address and IP address for a given subnet.

Interconnection of a Cisco DFA fabric to external networks is achieved through specified leaf switches operating as border leafs that peer with external standard unicast and multicast routing protocols. Advanced protocols such as Cisco Locator/ID Separation Protocol (LISP), Multiprotocol Label Switching (MPLS), Virtual Private LAN Services (VPLS), and Cisco Overlay Transport Virtualization (OTV) are all supported by the Cisco Nexus 7000 Series as well as the Cisco Nexus 7700 Series and can be connected to the border leaf for extending the Cisco DFA fabric or adding additional functionality.

The Cisco Nexus 7000 Series, Cisco Nexus 7700 Series, and Cisco Nexus F3-Series I/O modules offer a comprehensive solution to combine the border leaf as well as advanced protocol support within a single switch.

**Virtual Fabrics**

Virtual fabrics offer logical fabric isolation and segmentation within the fabric, extending the boundaries of segmented environments to different routing and switching instances. All these technologies can be combined to support host, cloud, and multitenancy environments.

Cisco DFA with its virtual fabrics supports VLAN scaling beyond the 4000-VLAN space traditionally offered by the 12 bits IEEE 802.1Q header in an Ethernet frame (Figure 2). By extending the traditional header value within the fabric, Cisco DFA has a 24 bit name space to uniquely address routing instances (VRF) and fabric global bridge domains (logical Layer 2 domains such as VLANs). This extended name space is achieved by Segment-ID.

**Figure 2. Segment-ID**
This extended name space removes the requirement for global VLANs and allows them to be locally significant to a leaf switch or a group of leaf switches. The stitching of local VLANs with the global Segment-ID allows the extension of Layer 2 domains across the fabric in conjunction with enhanced forwarding and routing within and between subnets (Figure 3).

**Note:** The Segment-ID solution consists in using two IEEE 802.1Q tag back-to back with a leading tag (outer tag) and a trailing tag (inner tag) for a total address space of 24 bits, allowing for the support of ~16M Layer-2 segments. Segment-ID is added/removed by the DFA Leaf nodes and is part of the Layer-2 Header.

A global VLAN only consists of a single IEEE 802.1Q tag, which is equally populated in the ethernet header as the Segment-IDs outer tag.

A non populated inner tag will decrease the amount of available Segment-ID, as only 12 instead of 24 bits are available.

**Example**

If global VLAN 1 is used, the Segment-ID 4'096 - 8'191 (0x1000-0x1FFF) cannot be used because of the non-populated inner tag.

If Global VLAN 1'000 is used, Segment-ID 4'096'000 - 4'100'095 (0x3E8000-0x3E8FFF) cannot be used because of the non-populated inner tag.

**Note:** Other than Nexus 5600, Nexus 6000 and Nexus 7000 with F3 Series I/O Module, the Nexus 5500 is not capable of using Segment-ID.

The Nexus 5500 can co-existence in a DFA fabric with the platform mentioned above but segments served by a Nexus 5500 require to use the classic VLAN space across all participating Switches. These VLANs are called “Global VLANs” and are global significant across the DFA Fabric.
Virtual fabrics use Layer 3 routing separation through Virtual Routing and Forwarding (VRF) to achieve separated routing instances for support of overlapping subnets in cloud and multitennancy environments. Each virtual fabric receives a dedicated routing identifier based on the extended 24 bit name space provided by the Cisco DFA Segment-ID. The routing separation across the fabric is achieved in a way similar to the way it is in an MPLS Layer 3 virtual private network (L3VPN), but using the Segment-ID to identify these instances uniquely and transport them across MP-BGP with the VPNv4/VPNv6 address-family (Figure 4).

The capability to configure virtual fabrics is exposed through the northbound open API of the Cisco Prime DCNM and can be used by an external orchestrator. The virtual fabric functions can also be configured manually through the Cisco Prime DCNM web interface. The Cisco DFA concept of applying configuration to the fabric at the time the workload requires it also applies to virtual fabrics; thus, preconfiguration of the whole fabric is not required. Fabric management provides advanced visibility into the virtual fabrics and offers a view of the member switches for a given virtual fabric whenever workload-related configuration is applied.
Cisco DFA Terminology

This section defines terminology specific to Cisco DFA.

Cisco DFA Fabric

To build a Cisco DFA fabric, a leaf switch must connect to a spine switch, another leaf switch, or a border leaf. In a typical folded Clos topology, a leaf switch or border leaf connects only to spine switches, but Cisco DFA also allows flexible topologies that vary from this model (Figure 5).

**Note:** Cisco DFA was extensively tested in a folded Clos topology (Leaf/Spine).

Figure 5. Clos Fabric

The Clos network was invented by Charles Clos in 1952 and represents the mathematically optimal architecture for networks, because every end node connected to the Clos fabric can always be reached by other devices through the same number of hops. This network fabric architecture enables the use of a horizontal scale-out model, which optimizes growth in a two folded way: the addition of spines increases the overall fabric bandwidth and resiliency while the addition of leafs increases the number of host-facing ports.

Cisco DFA Switch

Each switch participating in a Cisco DFA fabric is a Cisco DFA switch. Cisco DFA switches perform any of three roles: leaf switch, spine switch, or border leaf. These roles are each described in this section.

Cisco DFA Leaf

Cisco DFA leaf switches have ports connected to Classical Ethernet devices such as servers (host interfaces) and ports connected to the Cisco DFA fabric (fabric interfaces). Cisco DFA leaf switches can perform forwarding based on the enhanced control-plane functions of Cisco DFA optimized networking, which requires FabricPath with Segment-ID-based forwarding as well as the host/subnet control plane based on MP-BGP. The Cisco Nexus 5500 series can participate as a Layer 2-only leaf integrated into a Cisco DFA fabric, but it cannot perform Segment-ID-based forwarding, the distributed gateway functionality or run the host/subnet control plane.

Cisco Nexus 5500 Series switches can permanently coexist with DFA leaf switches in a given DFA fabric.

Cisco Nexus 5600 Series switches are DFA leaf switches which incorporate all existing DFA functionalities available on the Nexus 6000. The Cisco Nexus 5600 Series switches will support additional encapsulation for DFA...
(e.g. VXLAN) together with the Cisco Nexus 7000 Series or Cisco Nexus 7700 Series with the F3-Series I/O modules.

**Cisco DFA Border Leaf**
Cisco DFA border leaf switches are similar to Cisco DFA leaf switches in that they normally are physically attached to every spine switch in a standard folded Clos topology. However, a border leaf is used mainly to connect external network devices or services (firewalls, router ports, etc.) to a Cisco DFA fabric. Border leaf devices also can perform forwarding based on the enhanced control-plane functions of Cisco DFA optimized networking, which requires Segment-ID-based forwarding.

In addition, a border leaf learns external information from another fabric, WAN, or data center interconnect (DCI) connected to the Cisco DFA fabric and makes them available for the other switches within. This feature makes the border leaf a termination point for external unicast and multicast protocols, where it can serve as a translator between the Cisco DFA fabric and the rest of the world.

**Cisco DFA Spine**
Each Cisco DFA spine switch interconnects to all leaf and border leaf switches in the standard Clos topology. The pure spine switch role is defined by a Cisco DFA switch that has no servers, hosts, or other end nodes attached to it. Spine devices forward transparently based on Cisco DFA optimized networking with enhanced or traditional forwarding. Cisco DFA spine switches have a more static configuration, which is not modified when tenants or networks are added to the Cisco DFA fabric.

**Cisco DFA Route Reflector**
A Cisco DFA route reflector can exist on any Cisco DFA switch but is on the Cisco DFA spine switches by default (default templates). The purpose of the route reflector is the same as in classic BGP, in which it helps distribute a single routing update or withdraw a message to a large number of neighbors without the need to maintain a fully meshed BGP neighbor configuration and session relationship. The Cisco DFA route reflector is used to scale the Cisco DFA fabric efficiently.

**Note:** A typical DFA deployment should exist out of at least two DFA route reflector for redundancy and fault-tolerance purposes.

**Fabric Interface**
Fabric interfaces are the ports on a Cisco DFA switch that attach to other Cisco DFA switches: in other words, the fabric facing interfaces (Figure 6).
Cisco DFA fabric interfaces are Layer 2 Interfaces configured with Cisco FabricPath frame encapsulation. The modular architecture of Cisco DFA allows the encapsulation used on the fabric to be changed to a different frame encapsulation in the future with no need to change the architecture itself.

**Host Interface**

To transport data traffic across a Cisco DFA fabric and use network autoconfiguration, a leaf switch must receive traffic with intact IEEE 802.1Q headers to identify the required VLANs. The Layer 2 interfaces to which servers and other end nodes are attached are called Cisco DFA host interfaces (Figure 7).

**Fabric Management Network**

The fabric management network in Cisco DFA is a dedicated out-of-band (OOB) network responsible for bootstrapping and managing the individual Cisco DFA switches (spine, leaf, and border leaf) controlled by fabric management (Figure 8). The fabric management network is responsible for transporting the protocols required for the various fabric management functions, listed in Table 2.
Figure 8. Fabric Management Network

**Note:** Every Cisco DFA switch needs to connect through the Ethernet OOB port (mgmt0) to the fabric management network. Cisco DFA also supports in-band management and Power-on-Auto-Provisioning (POAP). The in-band version of operation has not been described in this document.

**Note:** Access to each switch console connection for management is recommended, but not required, for Cisco DFA.

Table 2. Functions and Protocols for Fabric Management

<table>
<thead>
<tr>
<th>Function</th>
<th>Protocol</th>
</tr>
</thead>
<tbody>
<tr>
<td>Device autoconfiguration (POAP)</td>
<td>DHCP, TFTP, and SCP</td>
</tr>
<tr>
<td>Fabric discovery</td>
<td>SNMP</td>
</tr>
<tr>
<td>Human-to-machine and machine-to-machine communication</td>
<td>XMPP</td>
</tr>
<tr>
<td>Automated network provisioning</td>
<td>LDAP</td>
</tr>
</tbody>
</table>

**Note:** So that POAP does not interfere with other DHCP servers, you should use a dedicated VLAN and subnet for the fabric management network. Cisco Prime DCNM has an interface for the Cisco Prime DCNM Access Interface (eth0) and for the Ethernet OOB port (eth1) of each Cisco DFA switch (for example, mgmt0). You have the option of interconnecting the fabric management network with your existing OOB management network.

**Cisco Prime DCNM Access Network**

The Cisco Prime DCNM access network is a separate segment on the network where a network administrator can connect to operation tools such an element manager or NMS. This network is provisioned as a subnet and interface on Cisco Prime DCNM separate from the fabric management network. The Cisco Prime DCNM access network for Cisco DFA is the network administrator-facing interface for accessing fabric management and is also used for connecting northbound APIs to orchestrators (Figure 9).
Fabric Management
Fabric management, often called the central point of management, or CPOM, is a ready-to-use management system based on Cisco Prime DCNM. It provides fabric management, visibility into the fabric, and an extensible set of functions to more efficiently control the data center fabric. Fabric management combines ease of deployment and use with standards-based control protocols and an extensive customization capabilities and integration with an operations support system (OSS).

The following capabilities are provided through the Cisco Prime DCNM functions:

- Device autoconfiguration is the process of bringing up a Cisco DFA fabric with preset configuration templates applied to any device joining the fabric, installing an image, or applying basic configuration.
- Cable-plan consistency checks the physical connectivity of the fabric against a documented cable plan for compliance. Lack of compliance prevents specific links from being active, thus protecting the fabric from errors.
- The common point of fabric access allows administrators to interact with the fabric as a single entity (system), instead of using a switch-by-switch approach, simplifying querying and troubleshooting.
- Automated network provisioning provides a new layer of automation integration in which the data center fabric switching infrastructure can be provisioned automatically and as a result of physical and virtual workload instantiation.
- Network, virtual fabric, and host visibility is provided by the management GUI and supports views of the fabric that display a single set of active network elements belonging to that group.

For more information about the installation of the Cisco Prime DCNM open virtual appliance (OVA) as well as about fabric management, see the following documents which can be found at http://www.cisco.com/go/dfa:

- Cisco Prime DCNM 7.0 OVA Installation Guide
- Cisco Prime DCNM Configuration Guide
- Simplify Fabric Management white paper

Cisco DFA Deployment Considerations
This section discusses DFA deployment considerations:

- Deployment preparation for the Cisco Prime DCNM OVA for VMware vSphere
- Cisco Prime DCNM features for Cisco DFA
Fabric deployment with integrated POAP workflow (templates, DHCP, image and configuration repository, and POAP definition)

Additional considerations according to Cisco DFA virtual fabrics, control plane, forwarding plane, and workload mobility, please see http://www.cisco.com/go/dfa.

**Deployment Preparation for Cisco Prime DCNM OVA for VMware vSphere**

The “Cisco Prime DCNM 7.0 OVA Installation Guide” provides comprehensive guidance about how to deploy the Cisco Prime DCNM OVA in a VMware vSphere environment. The “Simplify Fabric Management” white paper also provides additional guidance. The discussion here provides some additional notes to help you deploy the Cisco Prime DCNM OVA.

**Note:** Verify that your system includes the prerequisites for scaling and software versioning. These requirements can be found in the respective versions of the Cisco Prime DCNM release notes.

Fabric management connectivity for Cisco DFA must be through the OOB device management interface (mgmt0) of each Cisco DFA switch. The management interfaces for any given switch that supports Cisco DFA are connected to the same management subnet, which is also connected to the Cisco Prime DCNM fabric management interface. (aka. the eth1 in Figure 9)

During POAP operations, any nonconfigured switch makes a locally scoped DHCP request. The Cisco Prime DCNM DHCP server responds to these requests and assigns a temporary IP address, which is used for the Cisco NX-OS Software image and configuration download.

After POAP bootstrapping is complete and the Cisco NX-OS image and the startup configuration are installed, the switch boots with the given Cisco NX-OS version and configuration.

**Note:** The POAP process is a day-zero option for installing unconfigured switches, either with Cisco DFA or any other predefined configuration. POAP can also be used for disruptive software upgrades or replacement of a defective switch with the same software and configuration as long as any manual command-line interface (CLI) configuration changes to the switch are updated in the configuration repository in Cisco Prime DCNM.

Before deploying the Cisco Prime DCNM OVA, the hypervisor virtual switch (vSwitch) configuration must be prepared in the VMware vSphere environment. This switch can be a standard vSwitch, a distributed vSwitch, or a Cisco Nexus 1000V (Figure 10).

**Figure 10. vSwitch Configuration Example**
Prepare the two needed port groups or port profiles and have either dedicated physical uplinks or dedicated VLANs for the Cisco Prime DCNM virtual appliance, as shown in Figure 10. Make sure that the switch’s OOB network interface (mgmt0) and the Cisco Prime DCNM internal interface are connected to the fabric management network.

**Note:** It is recommended to separate the two DCNM Network from each other, even if they can exist on the same network segment or subnet. The usage of the two separate DCNM Network is always required, even when using it in the same Subnet.

For additional Information and guidance on Cisco Prime DCNM OVA installation, see the “Cisco Prime DCNM 7.0 OVA Installation Guide” and the “Simplify Fabric Management” white paper. ([http://www.cisco.com/go/dfa](http://www.cisco.com/go/dfa))

**Cisco Prime DCNM Features for Cisco DFA**

This section provides a quick overview of the Cisco Prime DCNM features for Cisco DFA.

After you log in to the Cisco Prime DCNM web user interface, the Welcome screen provides easy access to licensing, POAP, performance data collection, and Cisco Prime DCNM documentation. Cisco Prime DCNM has most functions preinstalled and preconfigured through the OVA installation and so is ready to use, apart from installing the necessary licenses (Figure 11).

**Figure 11.** Cisco Prime DCNM Welcome Screen

---

**Note:** Performance data collection is part of the Cisco Prime DCNM Advanced license and is not specific to Cisco DFA. Performance Collection was available in Cisco Prim DCNM before the availability of DFA and stays separate.

A evaluation license for the Cisco Prime DCNM Advanced functions can be obtained from the Cisco website.

The Summary Dashboard provides health information, resource inventory, topology overview, and performance data collection (Figure 12).
**Figure 12.** Summary Dashboard

<table>
<thead>
<tr>
<th>Health</th>
<th>Inventory</th>
<th>Topology</th>
</tr>
</thead>
<tbody>
<tr>
<td>Problems</td>
<td>Solved No Status</td>
<td>Add New Node</td>
</tr>
<tr>
<td>Unmanaged</td>
<td>Active</td>
<td>Breaker</td>
</tr>
<tr>
<td>Switching</td>
<td>Inactive</td>
<td>Faulty</td>
</tr>
<tr>
<td>Warning</td>
<td>Normal</td>
<td>Not Installed</td>
</tr>
<tr>
<td>Info</td>
<td>Healthy</td>
<td>Not Installed</td>
</tr>
<tr>
<td>Debug</td>
<td>Critical</td>
<td>Not Installed</td>
</tr>
</tbody>
</table>

**Note:** Performance data collection can be enabled by choosing Admin > Collection, but it requires prediscovered and licensed LAN switches.

**Note:** Discovery of LAN switches is part of the integrated switch launch process.

The Dynamic Fabric Automation Dashboard provides a single view of the dynamic discovered switch and link status in a spine-and-leaf topology (Figure 13).

The colored circles that represent the switches use the classic traffic-light notation (red, yellow, and green). The number of alarms, if any, is listed within the circle.

If you select a single switch by clicking it, all other connected switches will remain in the current color, and nonconnected switches will turn gray. Additional information about the selected switch will be shown in the table on the left of the page.

The links between the colored circles can be made visible by clicking the Show Links check box at the top right of the topology. The color scheme for the links is similar to the color scheme for the discovered switches.

The meaning of the node and link colors is presented on the left side of the Dynamic Fabric Automation Dashboard.

The view into the virtual fabrics is also part of the Dynamic Fabric Automation Dashboard. The Organization/Partition menu (marked by red circle) provides a topology view of the configured leaf switches for the given virtual fabric.

**Note:** Spine switches do not have configuration settings specifically for a single virtual fabric and serve the whole network. Therefore, they are represented in any virtual fabric.
Figure 13. Dynamic Fabric Automation Dashboard

With the Dynamic Fabric Automation Dashboard search field, represented by a magnifying glass on the left side, you can search for switches, virtual servers, and physical servers.

**Note:** Searching for virtual and physical servers requires discovery with the Virtual Station Interface (VSI) Discovery and Configuration Protocol (VDP) as described in IEEE 802.1Qbg clause 41. With DFA, Cisco added a Org TLV to the VDP protocol to include the Virtual-Machine (VM) name.

Cisco Prime DCNM also provides Network, Compute, and Storage Dashboards.

The Network Dashboard is a representation of the discovered switch inventory, which can also be shown by choosing Inventory > Switch.

The Compute Dashboard provides a view into a virtual machine, the hypervisor on which it runs, and the network connectivity of the hypervisor server (Figure 14).
Figure 14. Compute Dashboard

The Compute Dashboard provides numerous counters for the network administrator if you connect VMware vSphere vCenter with Cisco Prime DCNM.

**Note:** Connecting VMware vCenter to Cisco Prime DCNM does not provide virtual server search capabilities. The collection of this information requires the Cisco Prime DCNM Advanced license.

**Note:** Connection to VMware vCenter can be performed in Cisco Prime DCNM by choosing Admin (General) > Datasource > VMware. This connection requires an administrative VMware vCenter account.

Fabric Deployment with Integrated POAP Workflow
Cisco DFA integrated POAP workflow (Figure 15) is designed to simplify day-zero setup for any kind of POAP-compatible switch, but especially for the hardware and software that Cisco DFA supports (see Table 1 earlier in this document).

In addition to the POAP workflow, Cisco Prime DCNM provides as set of predefined templates for day-zero provisioning of a Cisco DFA fabric. These templates include definitions for the Cisco DFA leaf and spine roles for all hardware that Cisco DFA supports (Cisco Nexus 5500, 5600, 6000,7000 and 7700 Series).

If you do not want to use the predefined templates as they are, you can use the integrated POAP template editor to alter the templates or create new templates. DFA templates enable application of new config commands which in turn enables Cisco DFA fabric functionality.
Figure 15. Cisco DFA POAP Template Workflow

The Cisco DFA spine and leaf POAP templates do share common sections and will be extended given additional Features and Functions supported with Cisco DFA.

The General and Out-of-Band Configuration section provides the minimal configuration that a switch requires to operate; common information includes the switch name, admin user and password, console speed, and timeout values and the out-of-band interface (mgmt0) configuration, including its gateway.

In the Fabric: Manageability and Cable Plan section, you define the IP address and DNS address of the LDAP and XMPP server and the related password. Your Cisco Prime DCNM will be the LDAP and XMPP server IP address of the fabric management interface.

**Note:** XMPP requires the use of fully qualified domain names (FQDNs). Use the same FQDN for the Cisco Prime DCNM access network as for the fabric management network. For the fabric management network, the switch will create a static host-to-IP address mapping in the switch configuration. For the Cisco Prime DCNM access network, use a DNS server and make sure that Cisco Prime DCNM can resolve names against it.

Each Cisco DFA switch can participate in one or multiple XMPP groups. XMPP groups are used to communicate with a group of switches sharing a common purpose. To use XMPP and its group feature, you should create a group for all Cisco DFA switches (*node-all*), a group for all Cisco DFA spines (*node-spine*), and a group for all Cisco DFA leaves (*node-leaf*). If you require additional groups that combine a subset of switches, you can add these groups to that field.

Enforcement of a cable plan on the switches requires an additional CLI command. When you enable cable-plan enforcement, each switch checks the installed cable plan against its discovered neighbors.

Neighbor-switch discovery for the cable plan and consistency check is performed with specific type-length-value (TLV) extensions of the Link Layer Discovery Protocol (LLDP).

The In-Band Configuration section focuses on the IPv4 and IPv6 reachability within the DFA Fabric, which is required to create MP-BGP reachability between the differed DFA switches.

The backbone VLAN is used for MP-BGP neighbor sessions to exchange the host-route information of the Cisco DFA enhanced control plane. For the backbone VLAN, you can use an enhanced Cisco FabricPath feature (fabric forwarding control-segment) that injects the backbone information persistently into Cisco FabricPath control.
protocol, namely IS-IS. enhancement improves the reachability for the backbone IP addresses and therefore increases the resiliency for the BGP neighbor sessions, because traditional Cisco FabricPath learning is no longer required for the backbone-specific MAC addresses (Figure 16), flood due ARP or ND discovery (flood & learn) can be completely elimination on the backbone VLAN.

The backbone IP addresses can be overlapping, because the backbone information is isolated in a dedicated VRF instance (default VRF). This is true as long as host-facing network subnets do not exist in the default VRF. For the backbone VLAN, such isolation does not exist, and therefore the VLAN number must be chosen carefully.

**Note:** It is recommended to not use VLAN 1 as the backbone VLAN, as every default configured port do exist in this VLAN. In case a host will be connected to a such configured port, flooding can be introduced and/or the control-plane network can be compromised through external access.

**Note:** The DFA backbone VLAN is a global VLAN, which means it is global significant to the DFA fabric and not mapped to a Segment-ID.

The Segment-ID consists of using two IEEE 802.1Q tag back-to-back with a leading tag (outer tag) and a trailing tag (inner tag). A global VLAN only consists of a single IEEE 802.1Q tag, which is equally populated in the ethernet header as the Segment-IDs outer tag.

![Segment-ID](image)

A non populated inner tag will decrease the amount of available Segment-ID, as only 12 instead of 24 bits are available.

**Example**

If global VLAN 1 is used, the Segment-ID 4’096 - 8’191 (0x1000-0x1FFF) cannot be used because of the non-populated inner tag.

If Global VLAN 1’000 is used, Segment-ID 4’096’000 - 4’100’095 (0x3E8000-(0x3E8FFF) cannot be used because of the non-populated inner tag.

**Figure 16.** Cisco DFA Backbone and BGP Neighbors
In the Fabric: Layer 2 Control Plane section, the Cisco FabricPath switch ID is required to uniquely address the various switches in the Cisco FabricPath topology.

Cisco DFA has a set of defined switch roles: spine, leaf, and border (border leaf). This field is prefilled in the Cisco DFA POAP templates to assign the relevant role to a given switch.

The fabric forwarding identifier is used to inject a site-of-origin ID into the BGP prefixes. The fabric forwarding identifier uniquely identifies the site from which a router has learned a route. BGP can use the site-of-origin value associated with a route to prevent routing loops.

The Cisco FabricPath IS-IS control plane section maintains the topology and reachability between the switches. The MP-BGP base Fabric: Layer 3 Control Plane section is used to distribute host route objects among the leaves. To increase the efficiency of BGP configuration in a Cisco DFA fabric, a redundant route reflector configuration is prepared in the predefined Cisco DFA templates.

The BGP client subnet is used for the BGP neighbor configuration between the route reflector and the Cisco DFA leaves. Creation of a dedicated BGP neighbor statement for every leaf does not scale from a configuration perspective; therefore, a single BGP neighbor configuration, including a network ID and netmask prefix (neighbor \texttt{x.x.x.x/xx} remote-as \texttt{AS#}), is used. With the transport of the backbone information (MAC/IP and MAC/Switch binding) within Cisco FabricPath control-plane protocol (IS-IS) (the fabric forwarding control segment), the BGP neighbor session will be set up as soon as the Address Resolution Protocol (ARP) table for the neighboring route reflector and leaf is populated.

\textbf{Note:} Cisco FabricPath control-plane protocol, IS-IS, transports and installs static ARPs and static MAC entries for the backbone control SVI (Backbone VLAN) of all neighboring switches.

To simplify the import of a default-route (0.0.0.0/0) or other external routes to the Cisco DFA fabric, a common import route target is created on every leaf. This single import route target is used in any VRF instance; the related export route target, for example, on a border leaf, announces this information through BGP in the respective routing tables.

In the Distributed Gateway and Host Mobility section of the Cisco DFA leaf templates, network autoconfiguration and host mobility information is entered.

The anycast gateway MAC address is used for the distributed gateway functions. The global command sets this common MAC address and uses it across all VLANs on a given leaf. To help ensure virtual machine mobility between leaves, this value should be the same on every leaf.

A mobility domain defines a unique Layer 2 name space (for example, the VLAN domain and range), which can be represented by a virtual machine manager or data center definition (Figure 17).

A given leaf can belong to only one mobility domain, but within a Cisco DFA fabric, multiple mobility domains with overlapping VLAN ID spaces can exist. Overlapping VLANs from different mobility domains are uniquely mapped to Segment-IDs within the Cisco DFA fabric. This mapping offers the option to still use VLANs southbound to the servers and to map these VLANs to Segment-IDs northbound to the fabric.
For the dynamic assignment of VLANs in VDP-based network autoconfiguration, a range of dynamically assigned VLANs is required. VLANs from the system dynamic VLAN range are assigned using the next free one in the range. As soon as the VDP-enabled virtual switch signals a new Segment-ID, the leaf chooses the next free VLAN and sends this information back to the virtual switch. This VLAN is used to encapsulate communication from a specific Segment-ID.

**Note:** The usage of a IEEE 802.1Q tag for identifying a particular segment on the link between the Virtual-Switch and its associated attached leaf, has only link-local significance.

The second dynamic VLAN range, the core dynamic VLAN, is used for the VRF VLAN-to-Segment-ID mapping. This range is also assigned using the next free VLAN and is used only to map a Segment-ID to a VRF instance for unique routing-table identification. The core-vlan range is used to automatically auto-configure a vrf with the associated layer-3 or vrf Segment-ID where the corresponding core-vlan SVI interface serves as the egress interface for routed packets in the vrf that are sent toward the core or the DFA fabric.

**Note:** The "system fabric dynamic-vlans" is the super range of VLANs made available to DFA for dynamic mapping of Host as well as Core-VLANs to Segment-ID. The subsidiary VLAN range of "system fabric core-vlans" is used for dynamically allocating VLANs for the VRF isolation, which will end as Segment-ID from a fabric facing perspective.

**Important:** Do not use any VLAN out of the "system fabric core-vlans" on a front facing port!

**Example:**

```
system fabric dynamic-vlans 2500-3500
system fabric core-vlans 2500-2999
```

In the above example, you will start using VLAN 3000 and above for Layer-2 VLAN to Segment-ID mapping. Every VLAN from 2500 to 2999 will be used for the Layer-3 (VRF) to Segment-ID mapping.

Depending on the need for L2 VLAN or L3 VRF, the ranges need to be adjusted.
The Enhanced Virtual PortChannel (VPC+) Domain Configuration section incorporates all required steps to build a VPC+ domain including the identifiers for the vPC domain, Cisco FabricPath switch identifier, the Peer-Link PortChannel with its physical interfaces as well as the vPC peer-keepalive configuration.

**Note:** The Enhanced Virtual PortChannel (VPC+) section is not specific to DFA. Some assumption have been made within the POAP Templates to configure the VPC+ Domain.

**Note:** During the process of entering the serial numbers for the POAP definition, you are able to group the serial numbers in pairs of two with surrounding brackets. By using this notation, you can enter multiple comma separated IP addresses in the “VPC Peer IP” field, to assign the vPC peer-keepalive neighbor IP for a multiple of two switches. It is required to respect the order of the entered serial number to assign the correct vPC peer-keepalive neighbor IP address.

Additional configuration of the predefined Cisco DFA POAP templates is performed in the Interface, PortChannel, and Fabric Extender Configuration section (Figure 18). To reduce the number of POAP definitions that you need to create, you should use the same set of ports for fabric connectivity on the given spine and leaves. This approach helps ensure that a given POAP definition can be reused over multiple switches with the same purpose in the same or even multiple Cisco DFA fabrics.

**Figure 18.** Fabric Interface

In the same section, you can configure PortChannels and various interfaces:

- **Host interface:** An interface on the switch-port mode trunk with no VLAN range restriction
- **Access port:** A pseudo-trunk through which the switch-port mode trunk is used but with a given VLAN ID for the native VLAN (untagged)
- **PortChannel host:** A PortChannel (mode on) with multiple interfaces from a single leaf to a single host (server)
- **Breakout cable:** A cable that uses a single 40-Gbps interface and changes its mode to 4 x 10 Gbps

You can also create fabric extender (Cisco Nexus 2000 Series Fabric Extender) connectivity within the predefined Cisco DFA POAP templates for single- and dual-connected fabric extenders.

**Note:** Active-active fabric extender connectivity requires that the vPC+ domain be created first.
In general, the following southbound connectivity options for fabric extenders or hosts (servers, firewalls, etc.) are supported in the predefined Cisco DFA POAP templates as well as in Cisco DFA itself:

Valid southbound connection models (Figure 19):

1. Single-homed server with single link to one Cisco DFA leaf
2. Single-homed server with PortChannel to one Cisco DFA leaf
3. Dual-homed server with active-standby link to two Cisco DFA leaves
4. Dual-homed server with vPC to two Cisco DFA leaf (vPC+ domain)

Figure 19. Valid Southbound Connectivity

Note: All options use either the access port or IEEE 802.1Q trunk.

Valid fabric extender connection models (Figure 20):

5. Single-homed fabric extender with single link to one Cisco DFA leaf (Called ST-FEX or Straight-through FEX)
6. Single-homed fabric extender with PortChannel to one Cisco DFA leaf (Called ST-FEX or Straight-through FEX)
7. Dual-homed fabric extender with vPC to two Cisco DFA leaf (vPC+ domain) (Called AA-FEX or Active-Active FEX)
8. eVPC with fabric extender active-active link and dual-homed server with active-standby link or PortChannel

Figure 20. Valid Fabric Extender Connectivity

Conclusion

Cisco DFA with its four building blocks provides simplification, optimization and automation in the areas of fabric management, workload automation, optimized networking and virtual fabrics. With the concept of bundled functions within the four building blocks, we allow independent deployment in a modular fashion, which follows the individual speed of adoption and integration.
Cisco DFA uses standard and well-known protocols and functions. All of them are enhanced and optimized for scale and resiliency.

With the usage of POAP together with Cisco Prime DCNM, the Day-0 configuration of a single switch or a full network infrastructure can be reduced from days to minutes.

Cisco Prime DCNM provides network as well as network-service abstraction by using config-profiles. The change of the operational model increases the speed and consistency of workload deployment and decommissioning.

With combining the best of Layer 2 and Layer 3 forwarding, Cisco DFA optimizes the need for flood and learn to a minimum, by integrating an additional control plane for host and subnet information exchange. Additional to the optimized forwarding and learning, we also follow the concept of separation of an IP address from its location, similar to LISP (Location/ID separation protocol). The distributed gateway of Cisco DFA does not only provide a default gateway closest to the host, it also reduces the failure domain to a single or a pair of switches (if using vPC+).

The usage of virtual fabrics within Cisco DFA provides a similar approach for traffic separation as MPLS L3VPN does but for Layer 2 as well as Layer 3 traffic. Instead of a MPLS encapsulation and labels, we leverage the IEEE 802.1Q header of classic Ethernet. In order to be able to scale beyond the today available 4096 VLAN (12 bits), we doubled the IEEE 802.1Q header (24 bits), which gives us a name space of 16 million identifiers called Segment-ID.

For More Information

For more information and additional documents, white paper, release notes as well as installation and configuration guides, please visit http://www.cisco.com/go/dfa.