Dell EMC Networking, a Comparison Between Stacking and Virtual Link Trunking (VLT)

A technical white paper

Abstract

A technical white paper comparing Stacking and Virtual Link Trunking (VLT). This whitepaper provides the essentials needed to understand and configure Stacking and VLT.

September 2019

Revisions

Date	Description	Author
September 2019	Initial release 1.0	Umair Usmani

Acknowledgements

This paper was produced by the members of the Dell EMC network engineering team.

Table of contents

1 Introduction

In this paper, we discuss and compare Stacking and Virtual Link Trunking (VLT) which allow us to physically connect several Dell EMC devices so that they appear as single unit to a third device. VLT connects two devices whereas stacking can connect more than two devices depending on the switch series. Both these techniques provide multipathing which allows the administrators to create redundancy by increasing bandwidth and providing active/active paths between devices.

Both these techniques aggregate two identical physical switches to form a single logical extended switch. In stacking the control plane is unified, whereas in VLT the dual physical units form together as a single logical unit, but the control and data plane of both switches remain isolated.

2 Overview of stacking

In this technique, network switches are connected to operate as single unit called a stack. Such configuration can be used to quickly increase the capacity of the network.

Stacking makes it easier for users to expand their network without introducing the complexity of managing multiple devices. Stackable switches can be added or removed from the stack without disturbing the overall performance of the stack. Even in scenarios where one link or unit fails, data transfer continues uninterrupted. These are some of the reasons that make stacking a flexible, effective and scalable solution to add network capacity.

The maximum number of units that can be added to a single stack depends on the switch series. For further information see the **Appendix.**

2.1 Stacking architecture

The stack elects the management units for the stack management.

- Stack master – The primary management unit, also called the master unit.
- Standby – Secondary management unit.
- Stack units – The remaining units in the stack, also called stack members
- Stack group – Each set of four 10G ports, or each individual 40G port correspond to a stack-group.

The master holds the control plane and the other units maintain a local copy of the forwarding databases. From the stack you can configure both the system-level features and interface-level features that apply to all stack members.

The master synchronizes the stack unit topology, stack running configuration, and logs. In the event of switch failure, inter-switch stacking link failure, switch insertion or switch removal, the standby replaces it as the new master, and the switch with next highest priority.

2.2 Features of stacking

- **Single IP management:** When connected multiple switches form a stack with larger port counts. The stack is managed as a single entity. As one of the switches, in the stack acts as a master, the entire stack is managed through the management interface. Web, CLI, SNMP of the stack master.
- **Master failover with transparent transition:** In the case of stack master failure, the standby unit assumes the stack master role. As soon as the stack master fails, the standby unit initializes the control plane and enables all other stacks units with current configurations.
- **Nonstop forwarding on stack:** The non-stop forwarding (NSF) feature allows the forwarding plane of stack units to continue to forward packets even when the control and management planes restart as a result of a power failure, hardware failure, or a software fault on the stack master, and allows the standby switch to quickly take over as master.
- **Hot add/delete and firmware synchronization:** Units can be added to and deleted from the stack without power-cycling the stack. The units that are to be added to the stack must be powered off before they are cabled into the stack to avoid election of a new master unit and a possible downgrade of the stack. When the newly-installed unit in a stack is powered on, the stack firmware synchronization feature is enabled. This feature automatically synchronizes the firmware version with the version running on the stack master. Synchronization can either cause an upgrade or a downgrade of the

firmware on the mismatched stack member. Once the firmware is synchronized on a member unit, the configuration on the member is updated to match the master switch. Also, when the startup configuration on the master switch is saved, it is automatically saved on the other members of the stack as well.

2.3 Election of stack master

The election or re-election of stack master takes place based on the following considerations:

- The switch is currently the stack master.
- The switch has the higher MAC address.
- A unit is selected as standby by the administrator, and a fail over action is manually initiated or occurs due to stack master failure.

When a switch is added to the stack, one of the following scenarios takes place regarding the management status of the new switch:

- If the stack master function is enabled on the switch but another stack master is already active, then the switch changes its configured stack master value to disabled.
- If the stack master function is unassigned and there is another stack master already in the system, the switch changes its configured stack master value to disabled.
- If the stack master function is enabled or unassigned and there is no other stack master in the system, then the new switch becomes a stack master.
- If the stack master function is disabled, the unit remains a non-stack master.

If the entire stack is power cycled, the switch that was the stack master before the reboot will remain the stack master after operations are resumed. The unit number on the switch is configured manually, but to avoid any conflict, following scenarios takes place whenever a unit is added to the stack.

- If the unit number is manually configured and there is no other device using that unit number in the stack, the switch then starts using that configured unit number.
- If the added switch does not have a unit number, then the switch sets its configured number to the lowest number in the stack.

2.4 Addition of a switch to the stack

When adding a new member to the stack ensure that only the stack cables and no network cables are connected. Each stack port configuration is saved in the member unit. If a new switch is added to a stack of switches that are already powered and running, and that already have a stack master, the newly elected switch becomes a stack member rather than a stack master. The stack master automatically upgrades the firmware on the newly added switches. If a firmware mismatch is detected, the newly added switch does not join the stack, but holds until it is upgraded to the same version as that of the master switch. After firmware synchronization finishes, the running configuration of the newly added unit is overwritten with the stack master configuration.

The stack port configuration is always stored on the local unit and may be pre-configured before the unit is added to the stack. If there is saved information on the stack master for the newly added unit, the stack master applies that configuration to the new switch. Otherwise, the stack master applies the default configuration to the new unit. Hot insertion of units into a stack is not supported. Never connect two functional powered-up stacks together.

2.5 Removal of the switch from the stack

Before removing any member from the stack ensure that the other members of the stack will not become isolated from the stack due to the removal. Also, ensure that the ring topology can form a communication path around the member must be removed.

Note that when removing a switch from the stack, disconnect all the links on the stack member. Also, statically re-route any traffic going through this unit. When a unit in a stack fails, the stack master removes the failed unit from the stack. The failed unit reboots with its original running-configuration. If the stack is configured in a ring topology, then the master stack automatically re-routes the traffic around the failed unit. If the stack is not configured in ring topology, the stack may split and the isolated units will then reboot and re-elect a new stack master. If a switch is removed and you plan to renumber the stack, issue a no-member unit command in stack configuration mode to delete the removed switch from the configured stack information.

2.6 Stacking standby

The standby unit may be pre-configured or automatically selected. If the stack master fails, the standby unit becomes the new stack master. If there is no switch in the stack that is configured as a standby unit, the software automatically selects a standby unit from among the existing stack. When the failed master switch resumes normal operation, it joins the stack as a member if the new stack master has already been elected. If the new stack master fails, the member unit then takes over as the master switch.

2.7 Failover roles

If the stack master fails or it is powered off, it is removed from stack topology.

The standby unit detects the loss of peering communication and takes ownership of stack management, switching itself from the standby role to the master role. The distrusted forwarding tables are retained during the failover, as is the stack MAC address. The lack of a standby unit triggers an election within the remaining units for a standby role.

After the former master switch recovers, despite having a higher priority or MAC address, it does not recover its master role but instead takes the next available role. View further details through `show redundancy` command.

2.8 Creating a stack

The stack elects a master and standby unit at bootup based on the **Unit Priority** or **MAC Address.** The unit priority is user configurable. The available range is from 1 – 14, with a higher value indicating a higher priority. By using the no stack-unit priority command, the priority can be set back to zero. The unit that is higher in priority is selected as the master unit and the unit that is lower in priority is set as the standby unit.

In case the priority is same on both master and standby unit, the unit with the higher MAC value becomes the master unit. The stack takes the MAC address of the master unit and retains it unless it is reloaded.

The device supports stacking in a ring or a daisy chain topology. Dell Networking recommends the ring topology when stacking the switches to provide redundant connectivity.

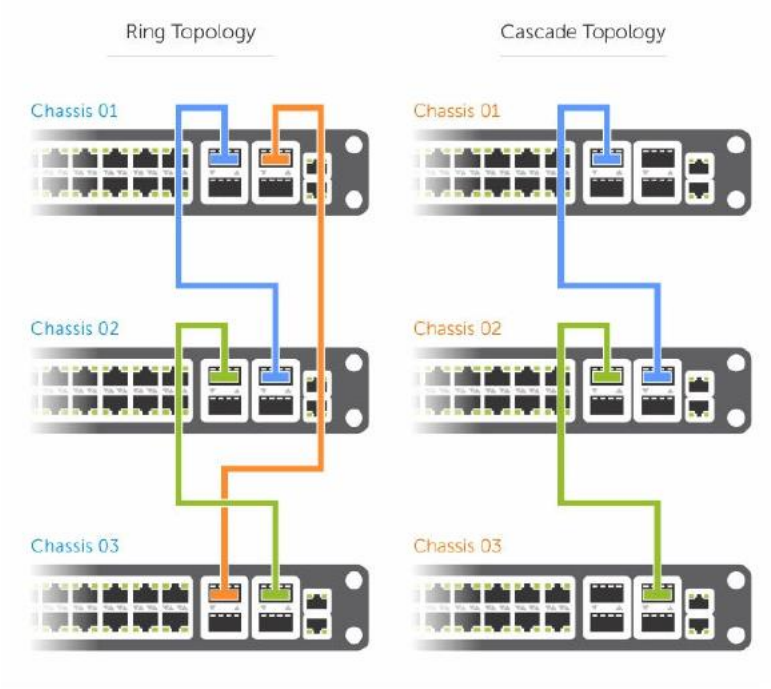
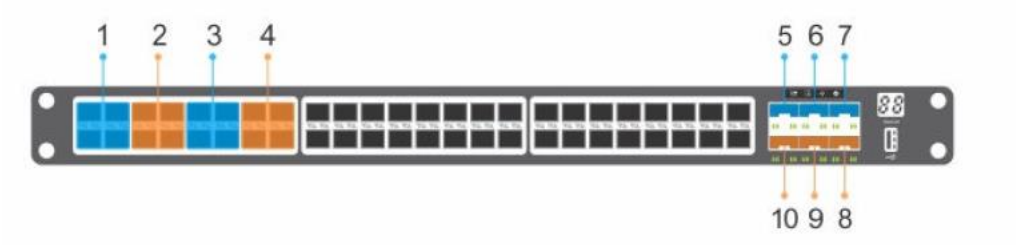


Figure 1: Ring and cascade topology for stack

In the case discussed here, we are using S4048 switch and stack groups 0 through 11 correspond to 10G stack groups with four ports each. Stack groups 12 to 17 are one 40G port each.



1. Stack group 0 (ports 1,2,3, and 4)
2. Stack group 1 (ports 5,6,7, and 8)
3. Stack group 2 (ports 9,10,11 and 12)
4. Stack group 3 (ports 13,14,15 and 16)
5. Stack group 12 (port 49)
6. Stack group 14 (port 51)
7. Stack group 16 (port 53)
8. Stack group 17 (port 54)
9. Stack group 15 (port 52)
10. Stack group 13 (port 50)

By default, each unit has stack unit number of 1 that can be changed using the `renumber` command.

```
StandbyUnit#stack-unit 1 renumber 2
 Renumbering management unit will reload the stack.
 Warning: Interface configuration for current unit will be lost!
 Proceed[confirm yes/no]:yes
 StandbyUnit#conf t
 StandbyUnit(conf)#stack-unit 2 priority 4
 MasterUnit(conf)#stack-unit 1 priority 10
```

The stacking units can be connected while they are powered up or down. When a switch is added to a stack, the management unit performs a system check on the new unit to ensure that the hardware type is compatible. Similarly, a check is performed on the Dell Networking Operating System Version.

The following is the sample configuration on one of the units:

```
MasterUnit(conf)#stack-unit 1 stack-group 1
MasterUnit(conf)#Mar 29 01:58:00: %STKUNIT1-M:CP %IFMGR-6-STACK_PORTS_ADDED: Ports Te 1/5
Te 1/6 Te 1/7 Te 1/8  have been configured as stacking ports. Please save and reset
stack-unit 1 for config to take effect
MasterUnit(conf)#exit
MasterUnit#Mar 29 01:58:48: %STKUNIT1-M:CP %SYS-5-CONFIG_I: Configured from  console
MasterUnit#write memory
!
Mar 29 01:58:56: %STKUNIT1-M:CP %FILEMGR-5-FILESAVED: Copied running-config to startup-
config in flash by default
MasterUnit#reload
```

The configuration looks as follows after it has been applied on both switches:

```
 MasterUnit>Show System

Stack MAC                  : e4:f0:04:3f:b9:15
Reload-Type                         : normal-reload [Next boot : normal-reload]

--  Unit 1 --
Unit Type                  : Management Unit
Status                     : online
Next Boot                  : online
Required Type              : S4048-ON - 54-port TE/FG (SK-ON)
Current Type               : S4048-ON - 54-port TE/FG (SK-ON)
Master priority            : 10
Hardware Rev               : 2.0
Num Ports                  : 72
Up Time                    : 6 min, 4 sec
Dell Networking OS Version : 9.10(0.1)
Jumbo Capable              : yes
POE Capable                : no
FIPS Mode                  : disabled
Burned In MAC              : e4:f0:04:3f:b9:15
No Of MACs                 : 3
[Output Omitted]
--  Unit 2 --
Unit Type                  : Standby Unit
Status                     : online
Next Boot                  : online
Required Type              : S4048-ON - 54-port TE/FG (SK-ON)
Current Type               : S4048-ON - 54-port TE/FG (SK-ON)
Master priority            : 4
Hardware Rev               : 2.0
Num Ports                  : 72
Up Time                    : 6 min, 1 sec
Dell Networking OS Version : 9.10(0.1)
```

```
Jumbo Capable              : yes
POE Capable                : no
FIPS Mode                  : disabled
Burned In MAC              : e4:f0:04:3f:ae:15
No Of MACs                 : 3
[Output Omitted]
```

3 Overview of Virtual Link Trunking (VLT)

VLT aggregates two identical physical switches to form a single logical extended switch. This single logical entity ensures high availability and high resilience for all its connected core switches, and clients. Even though both the switches form together as a single logical unit, the control and data plan of both switches remain discrete. As a result, we can apply a switch firmware upgrade without bringing down the network.

As high availability has become mandatory in modern data centers and enterprise networks, VLT plays a vital role connecting to all its access nodes with seamless traffic flow, efficient load balancing, and a loop-free mechanism.

3.1 VLT architecture

The VLT fabric consists of two nodes that provide a logical single switch view to the connected devices. However, each of the VLT peers has its own control and data planes and can be configured individually for port, protocol, and management behaviors.

The VLT application elects the primary node that is based on the lower MAC address. However, with the primary-priority command, the node with the least primary priority becomes the primary node. This election is not preempted, which means whenever there is a change in priority, the primary role does not change until the nodes are rebooted or the VLT process is restarted.

VLT-related information is shared between nodes through the specific reserved VLAN (VLAN 4094). The VLT database (VLT DB) is used to store the VLT control information to be exchanged between the VLT nodes. The local database (local DB) stores the MAC and ARP table entries.

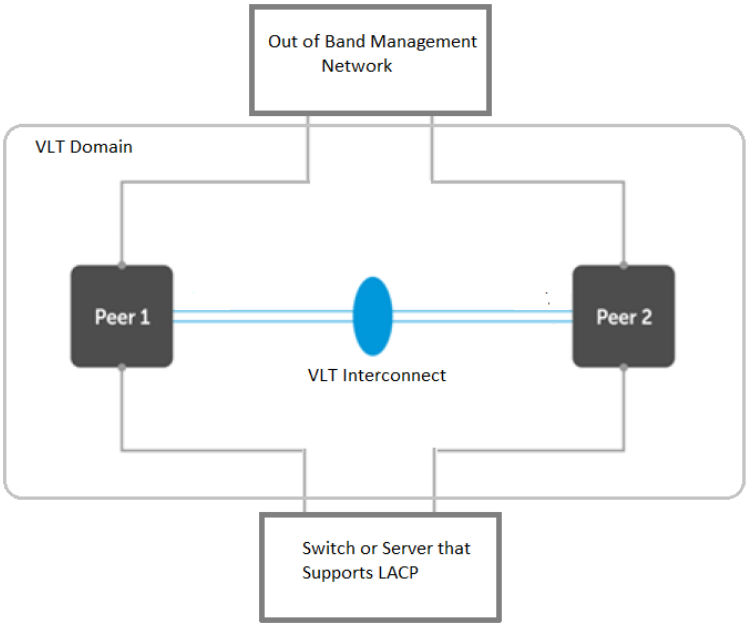


Figure 2: Virtual Link Trunking

3.2 VLT operation

Both the VLT nodes of domain always continue to forward data plane traffic in active/active mode. With the instantaneous synchronization of MAC and ARP entries, both the nodes remain active/active and continue to forward the data traffic seamlessly.

3.3 VLT functionality

The same VLT domain-id should be configured on both VLT nodes. The unit-id 1 and 2 for the nodes is configured automatically. For rapid convergence and optimal service, the same VLT MAC address should be configured on both the nodes using `vlt-mac` command. The priority of the primary node election is based on the lower system MAC-address of the switch, however with the primary-priority command, the VLT node with the least configured priority takes over as primary. This election will not be preempted. For example, when the primary node is reloaded, it is assigned the secondary role. The role change avoids disruptions in traffic flow due to the election process.

Election happens only during the initial configuration or when VLT is first launched. The VLT role election has no significance for the data traffic flowing through the VLT domain. It is only used for the control protocol exchange. VLAN ID 4094 is assigned automatically and internally reserved as a control VLAN for the exchange of VLT-related information between the nodes. The IPv6 address that is automatically assigned within the reserved range is mapped for VLAN 4094 for reachability between the VLT nodes.

For the VLT interconnect (VLTi) link, the discovery interfaces are configured on both the nodes, port-channel 1000 is automatically configured, mapping the physical discovery interfaces. The ports should be configured as no switchport from the default layer-2 mode while configuring the discovery interfaces.

For VLT port channels, the user should explicitly assign the vlt-port-channel ID to the configured port channel on both the nodes. This port channel identifier should be same across both the nodes.

3.4 VLT backup

VLT backup link is an additional link used to check the availability of the peer nodes in the VLT domain. When the VLTi interface goes down, the backup link helps to differentiate the VLTi link failure from peer node failure. If the VLTi link fails, all the VLT nodes exchange node liveliness information through the backup link.

3.5 Link failover scenarios in VLT

As shown in Figure 3, when the upstream layer-3 link 1 fails, the traffic is forced to take the alternate ECMP path to the VLT domain through link 2 to reach its destination.

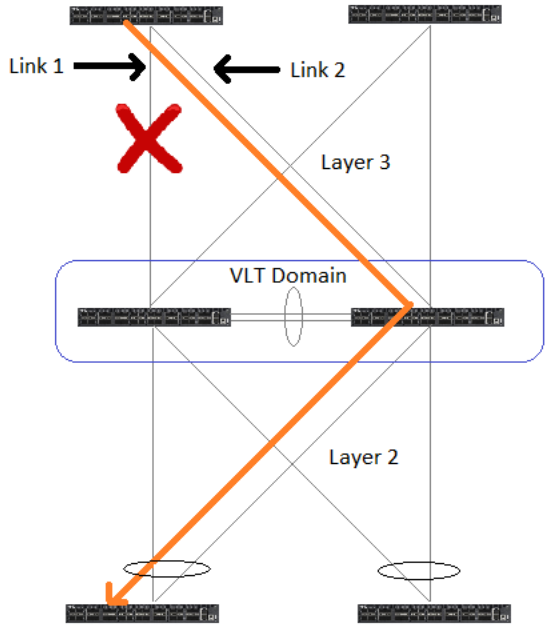


Figure 3: Link failure in upstream layer 3 link

In this scenario, where the link 3 in the VLT port channel fails as shown in Figure 4, the traffic will then pass through the VLTi and then take the link 4 as the MAC learned on the failed VLT is now mapped to the VLTi port.

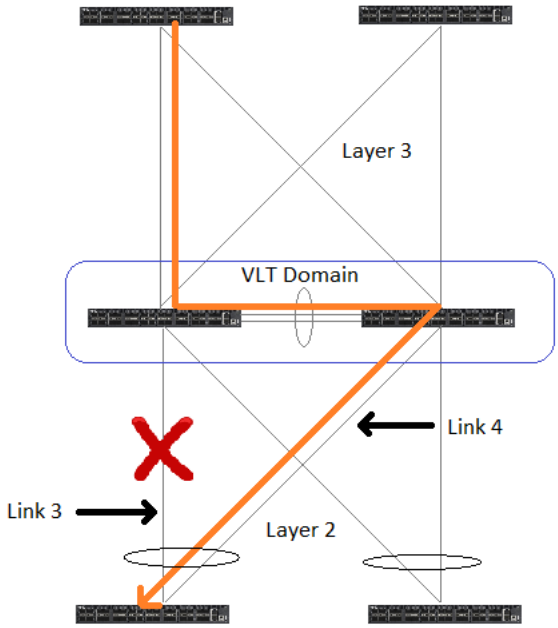


Figure 4: Traffic flow during link failure in the VLT port channel

The back-up heartbeat messages are exchanged between the VLT peers through the back-up links of the management network. Therefore, if the VLTi link fails and the peers continue to exchange the heartbeat messages, the primary VLT peer knows that the secondary VLT peer is up. Since the MAC/ARP entries cannot be synchronized between the two VLT peers, the secondary VLT node closes the VLT port-channel as shown in the Figure 5. Similarly, the north-south traffic only flows through the primary VLT peer. When the VLTi link is restored, the secondary peer waits for the pre-configured time for MAC/ARP entries to synchronize before passing the traffic.

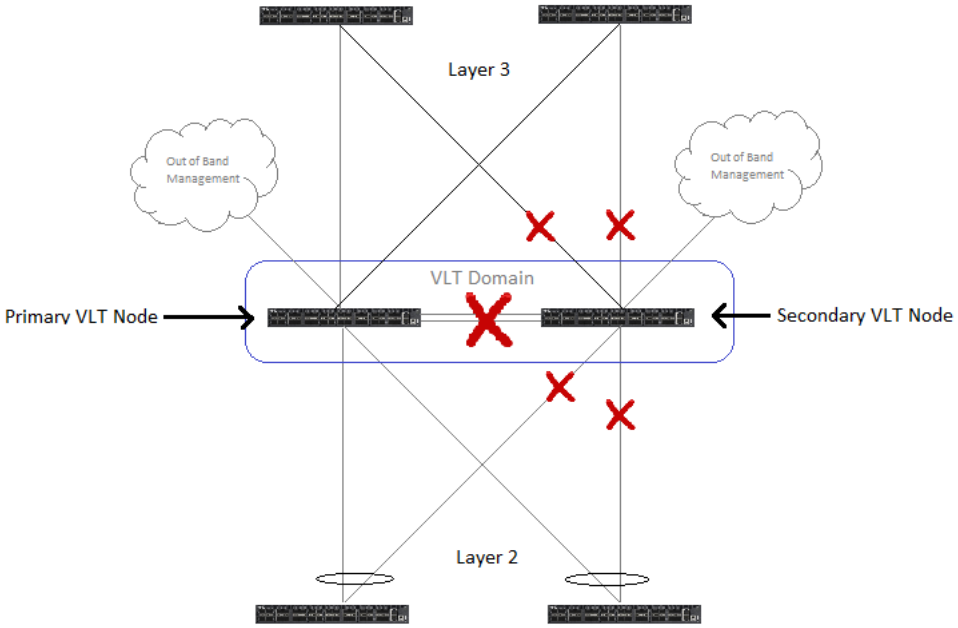


Figure 5: The VLTi failure (split brain scenario)

If the VLT peers don't even exchange the heartbeat messages, both nodes take the role of primary node and continue to pass the traffic.

3.6 VLT configuration

In order to configure VLT you must verify that both VLT peers are running the same operating system version.

To prevent loops in a VLT domain, enable STP globally using the `spanning-tree` mode command.

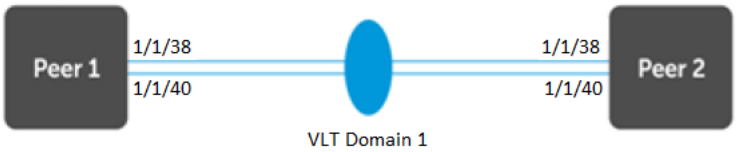


Figure 6: Lab setup for VLT configuration

VLT configurations on Peer 1

```
VLTPeer1(config)# spanning-tree mode rstp
VLTPeer1(config)# interface ethernet 1/1/40
VLTPeer1(conf-if-eth1/1/40)# no switchport
VLTPeer1(conf-if-eth1/1/40)# exit
```

```
VLTPeer1(config)# interface ethernet 1/1/38
VLTPeer1(conf-if-eth1/1/38)# no switchport
VLTPeer1(conf-if-eth1/1/38)# exit
VLTPeer1(config)# vlt-domain 1
VLTPeer1(conf-vlt-1)# discovery-interface ethernet 1/1/40
VLTPeer1(conf-vlt-1)# discovery-interface ethernet 1/1/38
VLTPeer1(conf-vlt-1)# vlt-mac 00:00:00:00:00:02
VLTPeer1(conf-vlt-1)# delay-restore 100
```

VLT configurations on Peer 2

```
VLTPeer2(config)# interface ethernet 1/1/40
VLTPeer2(conf-if-eth1/1/40)# no switchport
VLTPeer2(conf-if-eth1/1/40)# exit
VLTPeer2(config)# interface ethernet 1/1/38
VLTPeer2(conf-if-eth1/1/38)# no switchport
VLTPeer2(conf-if-eth1/1/38)# exit
VLTPeer2(config)# vlt-domain 1
VLTPeer2(conf-vlt-1)# discovery-interface ethernet 1/1/40
VLTPeer2(conf-vlt-1)# discovery-interface ethernet 1/1/38
VLTPeer1#
show vlt 1
Domain ID                  : 1
Unit ID                    : 1
Role                       : primary
Version                    : 2.0
Local System MAC address   : f4:8e:38:5f:47:ca
Role priority              : 32768
VLT MAC address            : 00:00:00:00:00:02
IP address                 : fda5:74c8:b79e:1::1
Delay-Restore timer        : 100 seconds
Peer-Routing               : Enabled
Peer-Routing-Timeout timer : 0 seconds
VLTi Link Status
    port-channel1000       : up

VLT Peer Unit ID    System MAC Address    Status    IP Address              Version
------------------------------------------------------------------------------------
  2                 f4:8e:38:56:2e:f6      up       fda5:74c8:b79e:1::2     2.0
VLTPeer2# show vlt 1
Domain ID                  : 1
Unit ID                    : 2
Role                       : secondary
Version                    : 2.0
Local System MAC address   : f4:8e:38:56:2e:f6
Role priority              : 32768
VLT MAC address            : f4:8e:38:5f:47:ca
IP address                 : fda5:74c8:b79e:1::2
Delay-Restore timer        : 90 seconds
Peer-Routing               : Enabled
Peer-Routing-Timeout timer : 0 seconds
VLTi Link Status
    port-channel1000       : up
VLT Peer Unit ID   System MAC Address    Status   IP Address              Version
------------------------------------------------------------------------------------
  1                 f4:8e:38:5f:47:ca      up       fda5:74c8:b79e:1::1     2.0
```

4 Conclusion

Both VLT and stacking technologies create a virtual switch where several switches are combined into one. This is not possible using standalone switches.

Even though both these technologies give us similar capabilities, stacked switches act as a single switch from both data plane and control plane, so if switches need to be updated, all the switches need to be rebooted, and the network fails. VLT offers one data plane but individual control planes, and thus each switch can be managed and upgraded separately without full network downtime. Some of the important points to remember when comparing VLT and Stacking are captured in the following table:

	OS	Advantages	Disadvantages
VLT	10, 9	• Allows the lifecycle management without any disruption to live traffic. • Provides the necessary link and device redundancy needed by any upstream or downstream connectivity. • Provides layer 2 and layer 3 connectivity independent of distance.	• VLT is limited to 2 peers.
Stacking	9	• Accommodates multiple switches depending upon the switch series. • The single logical chassis allows for easier maintenance as the stack is managed as one logical unit.	• Stack switches act as a single switch, so if a switch needs to be updated all the switches in the stack need to be rebooted.

5 Appendix

Campus switches

Campus switches run OS6, and the maximum number of units they can support in a stack is given in the table below. Note that N1500, N2000 and N3000 cannot be stacked together. In OS 6.6, N3000 doesn't support this branch of code.

Table 1 Number of switches per stack.

Switch series	Maximum number of switches in stack
N1500 series	4 units
N2000 series	12 units
N3000 series	8 units
N3000E-ON series	12 units and 8 units if they are stacked with N3000
N3132 series	12 units and 8 units if they are stacked with N3000
N3100 series	12 units
N3048	6 units

Data center switches

The OS9 for data center switches supports stacking. The maximum number of units that can be supported in a stack for given switch series are mentioned in the table below.

Table 2 Number of switches per stack

Switch series	Maximum number of switches in stack
S4048 series	6 units
S3100 series	12 units