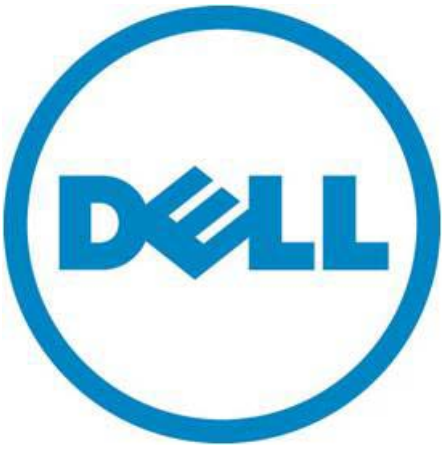
Stacking Dell PowerConnect 7000 Series Switches

A Dell Technical White Paper

www.dell.com • support.dell.com



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.

© 2011 Dell Inc. All rights reserved. Reproduction of this material in any manner whatsoever without the express written permission of Dell Inc. is strictly forbidden. For more information, contact Dell.

Dell, the *DELL* logo, and the *DELL* badge, *PowerConnect*, and *OpenManage* are trademarks of Dell Inc. Other trademarks and trade names may be used in this document to refer to either the entities claiming the marks and names or their products. Dell Inc. disclaims any proprietary interest in trademarks and trade names other than its own.

PC7024, PC7048, PC7024P, PC7048P, PC7024F, PC7048R, PC7048R-RA, PCM6348

July 2011

Contents

Introduction

This white paper explains the purpose and operation of the stacking feature in the Dell™ PowerConnect™ 7000 Series Gigabit Ethernet switches. The PowerConnect 7000 series is Dell's most advanced switching product line, offering advanced switching capabilities including high-density, high-performance stacking, and 10 Gigabit Ethernet capabilities that scale from the small business to the Enterprise Edge. Stacking allows multiple switching units to be combined together to act as a single, high-performance, highly resilient switching unit with a single management interface. Units can be added to increase throughput as needed. With each stack unit supporting up to 184 Gbps in switch capacity, the customer can have almost 2 terabits of capacity in a single stack.

In addition, stacks can be composed of an interchangeable mix of PowerConnect 7000 Series switches and PowerConnect M6348 switches, enabling administrators to continue to leverage the full utility of previous-generation switches while transitioning to the latest-generation equipment. See "Mixing M6348 and PC7000 Series Switches in a Stack" on page 5.

Applicability

This paper applies to the PowerConnect 7000 series switches, which includes the PC7024, PC7048, PC7024P, PC7048P, PC7024F, PC7048R, and PC7048R-RA Dell model numbers. Each PowerConnect switch has two bays that can be customized to support a stacking or an uplink configuration. Each bay can contain a CX-4, SFP+, or a 10GBase-T module. Stacking is supported only on CX-4 modules in either or both bays. CX-4 stacking modules can be configured in 16 Gbps stacking mode or 10 Gbps Ethernet uplink mode. CX-4 stacking modules default to stacking mode, where the maximum cable length is 3 meters. When the stacking modules are configured to operate in Ethernet mode, the maximum cable length is 12 meters.

The PCM6348 has two 10 Gbps SFP+ ports, and two separate 10 Gbps Ethernet ports that support CX-4 modules for stacking.

The following table summarizes the features of the PowerConnect 7000 series and PCM6348 switches.

Part Number	Gigabit Ethernet Ports (RJ-45)	Gigabit Ethernet Ports (SFP)	Modules (Stacking, SFP+, 10GBASE-T)	Other Features
PCM6348	16[1]	2	4	
PC7024	24	4[2]	2	
PC7024P	24	4[2]	2	PoE[4]
PC7024F	4	24	2	
PC7048	48	4[3]	2	
PC7048P	48	4[3]	2	PoE[4]
PC7048R/ PC7048R-RA	48	4[3]	2	Top-of-Rack[5]

1. The PCM6348 also has 32 internal ports that connect to server blades in a chassis.
2. Shared with ports 21–24.
3. Shared with ports 45–48.
4. Each copper port can provide up to 30W of power to an external powered device.
5. The difference between the PC7048R and PC7048R-RA switches is the airflow direction.

Stacking and Management

An important advantage of stacking is that it provides a consolidated interface for management of multiple switches when linked together. When a stack is already deployed in the network, operators can add units to the stack as their port requirements increase, with minimal administrative overhead

required for reconfiguration. Additional stack members can immediately utilize existing configuration information such as routing and switching configurations, VLANs, ACLs, port profiles, and security certificates.

Stacking and Performance

For situations where there is a need to pass traffic between switches and the aggregate bandwidth required between PowerConnect 7000 Series switches does not exceed 64 Gbps (2 ports, 16 Gbps Tx and Rx each), a stacking configuration offers an attractive alternative to Link Aggregation Groups (LAGs). Stacking configuration is generally transparent to the operator and does not require configuration beyond cabling. In addition, failover times are generally faster in a stack configuration than in spanning tree or link aggregation group configurations. Note that other PowerConnect Series switches may have different supported bandwidths for stacking.

Stacking and Redundancy

By connecting a cable from the last switch in a stack back to the first switch, the operator ensures that a stack has the protection of redundant paths for control and data traffic, including support for LAGs configured across multiple switches. This means that any single point of failure (a switch or a stack cable failure) will not affect the overall operation of the remaining stack elements.

Figure 1 shows two stacking configurations with a single stacking module on each switch. In the recommended configuration, a stacking cable connects each switch to the next, and the top switch is connected to the bottom switch to form a complete ring.

Figure 1. Stack Configured in a Ring

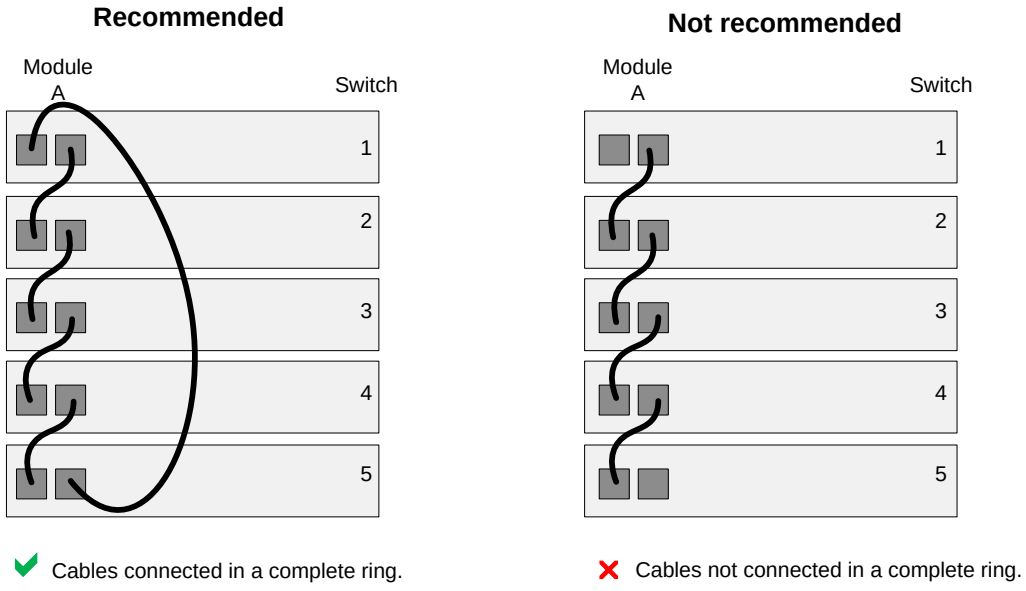
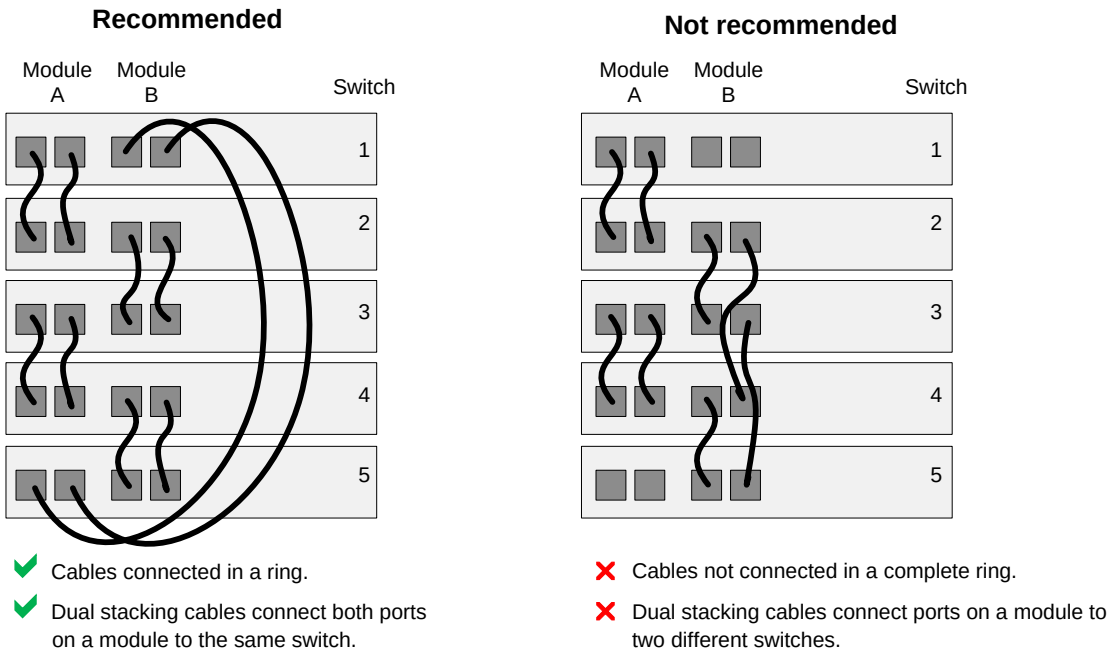


Figure 2 shows a recommended stacking configuration with dual stacking modules on each switch. Each switch connects to another with 128 Gbps of total bandwidth (4 ports @ 16 Gbps Tx and Rx each). Note that in the recommended configuration, the dual stacking cables from each module both connect to the same switch. In the not-recommended configuration, the dual stacking cables on switch 2, module B and switch 3, module B connect to two different switches.

NOTE: To ensure full bandwidth when using redundant links between two switches, be sure not to split the links across modules. Keep redundant links isolated to a single module on each end as shown in the recommended configuration in Figure 2.

Figure 2. Dual Stacking Modules Configured in a Ring



Mixing M6348 and PC7000 Series Switches in a Stack

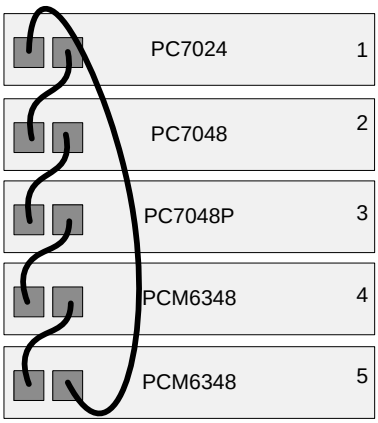
PC7000 series and M6348 can be used interchangeably in a stack of up to 12 units. As with PC7000-series only stacks, all switches in a mixed stack must have the same firmware version. Starting with the 4.1.0.6 software release, a single image supports both switches. No additional configuration is required.

Either the PC7000 series switch or the M6348 switch can be the master.

NOTE: A stack that includes an M6348 switch (as the master or a member) can operate only in Normal mode—not in Simple mode.

Figure 1 shows a stack with a mix of PC7000 series and M6348 switches.

Figure 3. Mixed Switches in a Stack



Power-Up Sequencing Considerations

One unit in the stack acts as the master unit. The master manages all the units in the stack. A second switch is elected as the standby unit, which becomes the master if the master unit is unavailable.

The administrator can manually configure which unit is elected as the standby, or the system can select the standby automatically. To configure it manually, the administrator can use the Stacking Management > Unit Configuration page in the web interface or the `standby` command in the CLI.

NOTE: The terms "master" and "manager/management unit" are used interchangeably throughout this document.

The selection of the manager and standby units is important in situations where the administrator wishes to utilize the serial port for management, perhaps as a backup console. Management of the stack via the out-of-band service port or the in-band ports is transparent in a stack.

In a stacking configuration, the power-up sequence determines the manager and standby units. The switch that the operator selects as the manager should be powered up first and should be allowed to fully come up before the other stack units are powered up. The standby switch should always be directly connected to the manager. Once the manager and standby have powered up fully, other

members of the stack can be powered on sequentially by powering up the switch adjacent to the last switch powered on.

A stack of units is managed and acts as a single entity when the units are connected together and are operational. If a unit cannot detect a stacking partner, the unit automatically operates as a stack of 1 with itself as the master. If a stacking partner is detected, the switch always operates in stacking mode.

When units are operating together as a stack, the following activities occur:

- All units are checked for firmware version consistency on startup. By default, units with older versions of firmware are automatically upgraded with the operational firmware version on the master switch as they join the stack. If a unit has a newer software version than the master, it is not downgraded to the master's version by default; however, the administrator can use the following command to enable this functionality:

  ```
  console#boot auto-copy-sw allow-downgrade
  ```

  Before adding a member that has newer version of software to a stack, the administrator should enable the automatic downgrade feature on the master switch.

- Switch management and protocols such as OSPF are active only on the master, but apply to all members of the stack. Unless administratively disabled, the Nonstop Forwarding (NSF) feature periodically checkpoints the running configuration and the application state between the master and standby switches during normal stacking operation. If the master fails, the standby switch takes over operation of the stack.
- Data forwarding is active on all units in the stack, including the master. Data forwarding continues to operate should the master become unavailable.

Initial Installation and Power-up of a Stack

Follow these instructions to create a stack:

NOTE: Install units in a rack whenever possible to prevent the units and cables from being disturbed.

1. Install all stacking cables. Fully connect all cables, including the redundant stack link.

CAUTION! We highly recommend that a redundant link be installed to provide stack resiliency.

2. Select a unit to be the manager unit. Power this unit up first.
3. Monitor the console port on the manager unit. The unit will automatically become a manager unit. If not, renumber the unit as desired. The Stack Management > Unit Configuration page in the web interface or the `switch renumber` command in the CLI can be used to change the unit number.
4. If desired, pre-configure other units to be added to the stack. This is not usually necessary.
5. Power on a second unit, making sure it is adjacent (directly connected) to the manager unit. This ensures the second unit comes up as a member of the stack, not as the manager of a separate stack.
6. Monitor the manager unit to see that the second unit joins the stack. Use the `show switch` command to determine when the unit joins the stack. It will be assigned a unit number (unit #2, if it has the default configuration). The output of the `show switch` command should indicate the unit status as “OK” if the member has been successfully added to the stack. If the unit status is "code mismatch" and stack auto-upgrade is disabled, then use the `copy image unit` command to update the code on the unit. If the unit status is "Cfg Mismatch", then use the `no member <unit-number>` command to resolve the issue.
7. If desired, renumber this stack unit using the `switch renumber` command.
8. Repeat steps **4** through **6** to add additional members to the stack. Always power on a unit directly connected to the units already in the stack.
9. Enter the `show switch stack-port counters` command on the manager and see if there are any stack port errors being reported. Replace the stacking cables/modules if stack errors are being reported and begin the power-up sequences again for the affected units and any downstream units.

Selecting the Master Unit

A stack manager is elected or re-elected based on the following considerations, in order:

- Whether the switch that was previously the stack manager.
- Whether the switch has the higher MAC address.

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

- If the switch was previously designated as the stack master but another master unit is already active, then the switch changes its configuration to be a slave unit.
- If the switch was not designated as the stack master and there is another stack master in the system, then the switch changes its configuration to be a slave unit.
- If the switch is enabled as the stack master or there is no other stack master, then the switch becomes stack master.
- If the switch is not enabled as the stack master, the unit remains a slave unit.

The administrator can manually set the unit number for the switch using the Stack Management > Unit Configuration page in the web interface, or the `switch renumber` command in the CLI. To avoid unit-number conflicts, one of the following scenarios takes place when a new member is added to the stack:

- If the switch has a unit number that is already in use, then the unit that is added to the stack changes its configured unit number to the lowest unassigned unit number.
- If the added switch does not have an assigned unit number, then the switch sets its configured unit number to the lowest unassigned unit number.
- If the unit number is configured and there are no other devices using the unit number, then the switch starts using the configured unit number.
- If the switch detects that the stack already has the maximum number of units, making it unable to assign a unit number, then the switch sets its unit number to “unassigned” and does not participate in the stack.

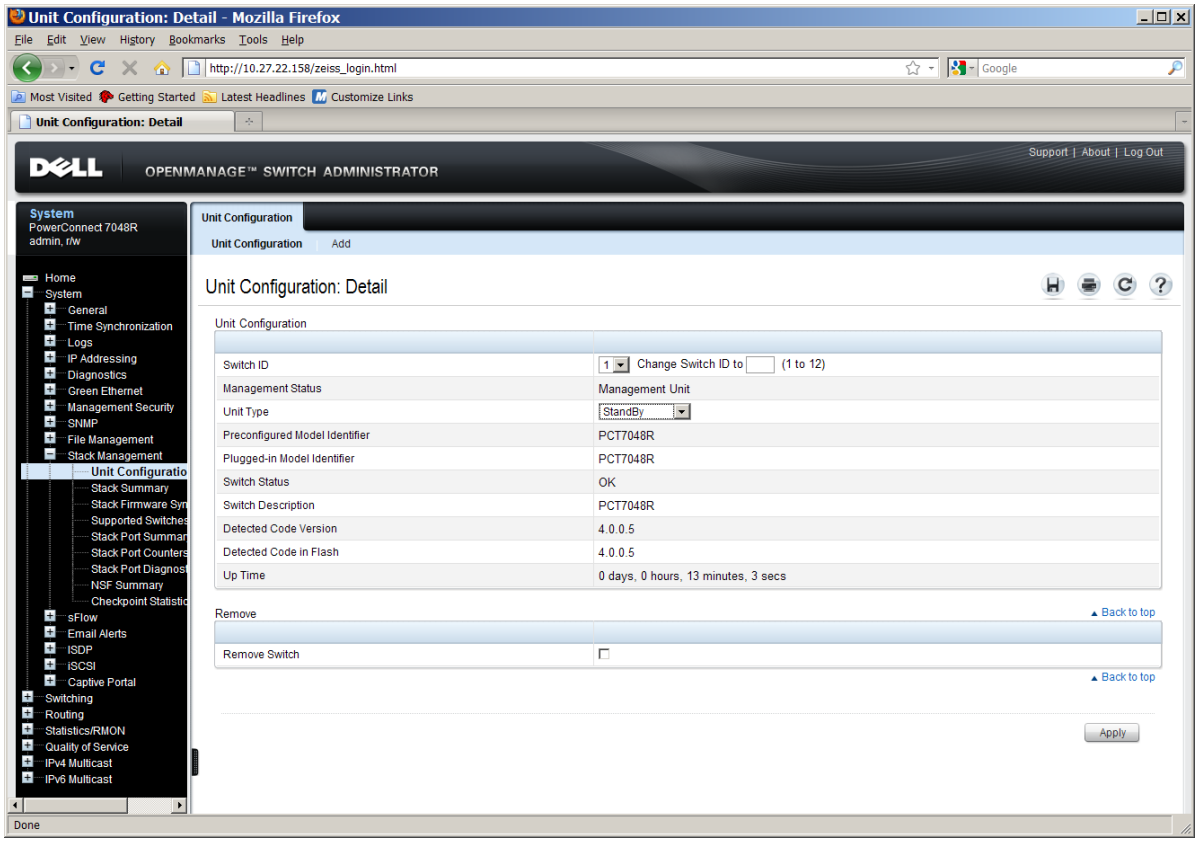
If a new switch is added to a stack of switches that are already powered and running and already have an elected master unit, the newly added switch becomes a stack member rather than the master. On the master unit, if there is no saved configuration for the newly added unit, it applies the default configuration. If there is a saved configuration on the master for the newly added unit, it would apply the saved configuration to the new unit. If the entire stack is powered OFF and ON again, the unit that was the master before the reboot will remain the master unit after the stack resumes operation.

Selecting the Standby Unit

When the stack is formed, one of the units is automatically selected as the standby unit for the stack. The standby unit takes over as manager if the current manager fails. Alternatively, the administrator can specify the standby unit.

To configure the standby unit using the CLI, use the `standby` command in Stack Configuration Mode.

To configure the standby unit via the Web interface, view the **Stacking > Unit Configuration** page and select **Standby** for the **Unit Type**.



Updating the Firmware on a Stack

Automatic Update

By default, firmware synchronization is performed automatically when a unit is powered up on a stack. Because firmware synchronization makes no attempt to check for the latest version of firmware, the following procedure is recommended for bringing all members of a stack onto the same version of code.

NOTE: Schedule some downtime as this process will reset the entire stack and affect all stack users.

1. Use the following commands to enable firmware synchronization, if needed, and write the running configuration to the saved configuration:

   ```
   console#boot auto-copy-sw
   console#copy nvram:config copy system:running-config nvram:startup-config
   ```

2. Load the more recent firmware image onto the stack master as the backup image.
3. Set the stack master to reboot from the new image:

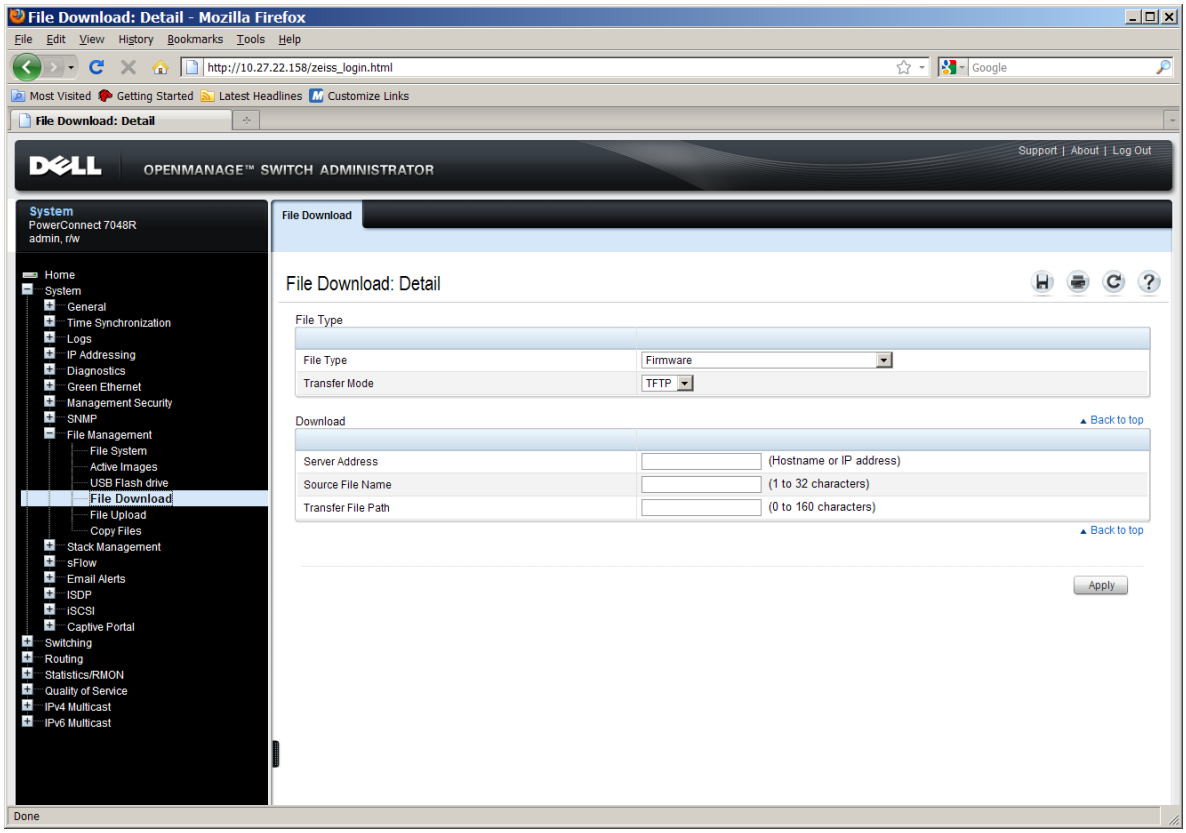
   ```
   console#boot system backup
   ```

4. Reload the stack command. This resets the entire stack. The switches stop forwarding during the reset.
5. When the stack has reloaded, check the running firmware version using the `show version` command. Check that all stack members have joined the stack using the `show stack` command. Members that failed to update may be recovered using the manual update procedure described in the following section.
6. If successful, disable firmware synchronization on the stack and save the running configuration.

Manual Update

The administrator can perform firmware updates on the stack by using the CLI or the Web interface. From the CLI, use the `copy ftp` command, which uses the FTP protocol for file transfer. To update the firmware by using the Web interface, use the options available on the **System > File Management > File Download from Server** page.

NOTE: When stacked, the PC7000 switches require that the same version of firmware be installed on every switch member.



When connected in stack, the `copy ftp` or `copy tftp` commands will distribute the downloaded image to all the connected units of the stack.

In the following output, the image with a file name of image.stk is downloaded from the FTP server with an IP address of 10.27.64.141.

```
console#copy ftp ftp://10.27.64.141/image.stk image user admin password
test1234
```

```
Mode..................................... FTP
FTP Server IP............................ 10.27.64.141
FTP Path.................................
FTP Filename............................. PC7000_4.0.0.6.stk
Data Type................................ Code
Destination Filename..................... image
```

Creating a Separate VLAN for File Downloads

When updating the firmware, it is helpful to keep the in-band management port in a different VLAN and configure the port VLAN ID (PVID) appropriately to avoid the possibility of network congestion or flooding issues impacting the file download.

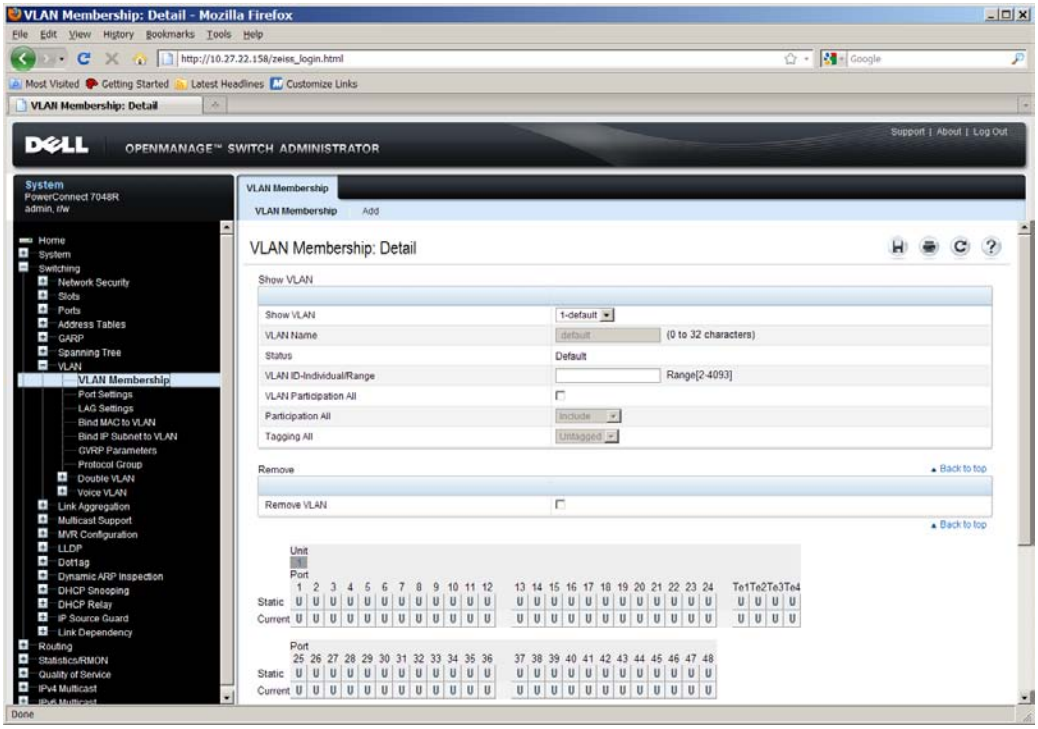
The CLI commands in the following example show how to configure port gi1/0/17 as an in-band management port for firmware downloads or management access.

```
console#configure
console (config)#vlan database
console (vlan)#vlan 1000
console (vlan)#exit
console (config)#interface vlan 1000
console (config-if-vlan1000)#ip address 192.168.21.11 255.255.255.0
console (config-if-vlan1000)#exit
console (Config)#interface ethernet gi1/0/17
console (config-if-gi1/0/7)#switchport mode general
console (config-if-gi1/0/7)#switchport general pvid 1000
console (config-if-gi1/0/7)#switchport general allowed vlan add 1000
console (config-if-gi1/0/7)#switchport general allowed vlan remove 1
console (config-if-gi1/0/7)#exit
console (config-macal)#management access-list MGMT_VLAN
console (config-macal)#permit ip-source 192.168.21.0 mask /24 vlan 1000
console (config-macal)#service ssh
console (config-macal)#exit
console (config)#management access-class MGMT_VLAN
```

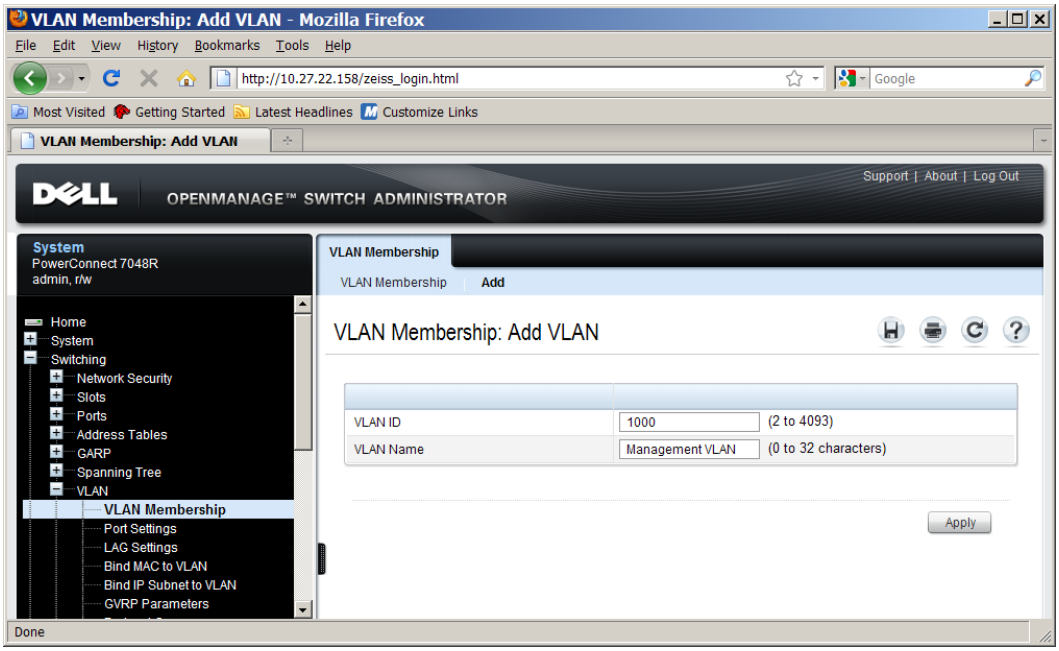
The switch now segregates traffic arriving on port gi1/0/17 onto VLAN 1000. All untagged packets that enter the port are tagged with a VLAN ID of 1000. Additionally, only hosts with an IP address in the 192.168.21.XXX subnet are allowed access to the switch using SSH. The 192.168.XXX.XXX address block is a private address space per RFC 1918. As an added security measure, network administrators can configure their organization's edge routers to drop ingress and egress traffic destined to this address block.

To perform the same configuration by using the Web interface, use the following steps:

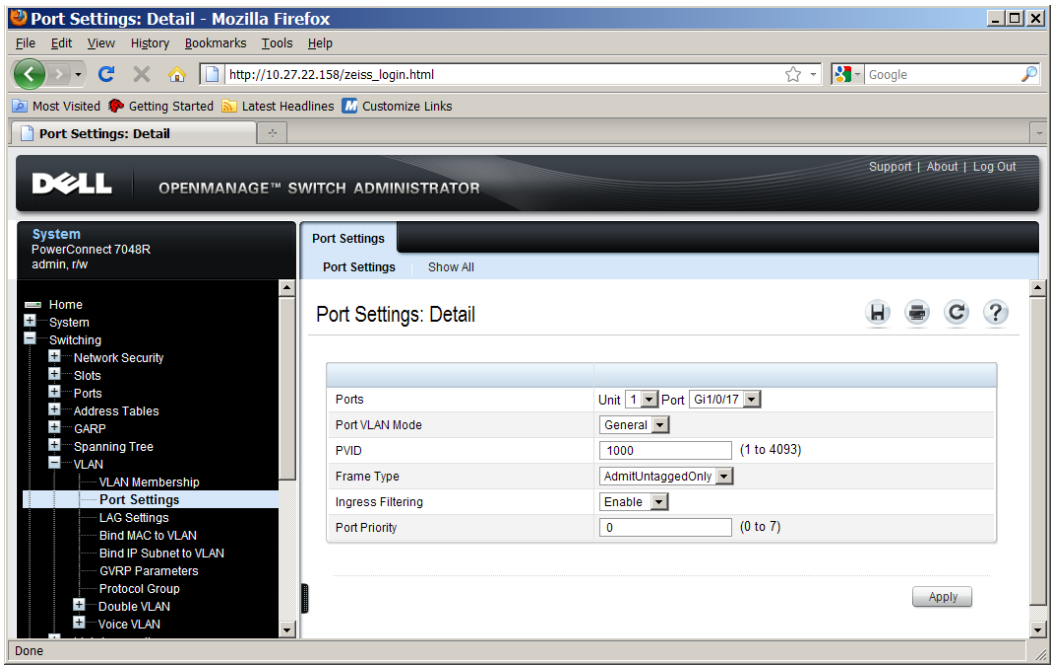
1. From the **Switching > VLAN > VLAN Membership** page, click **Add**.



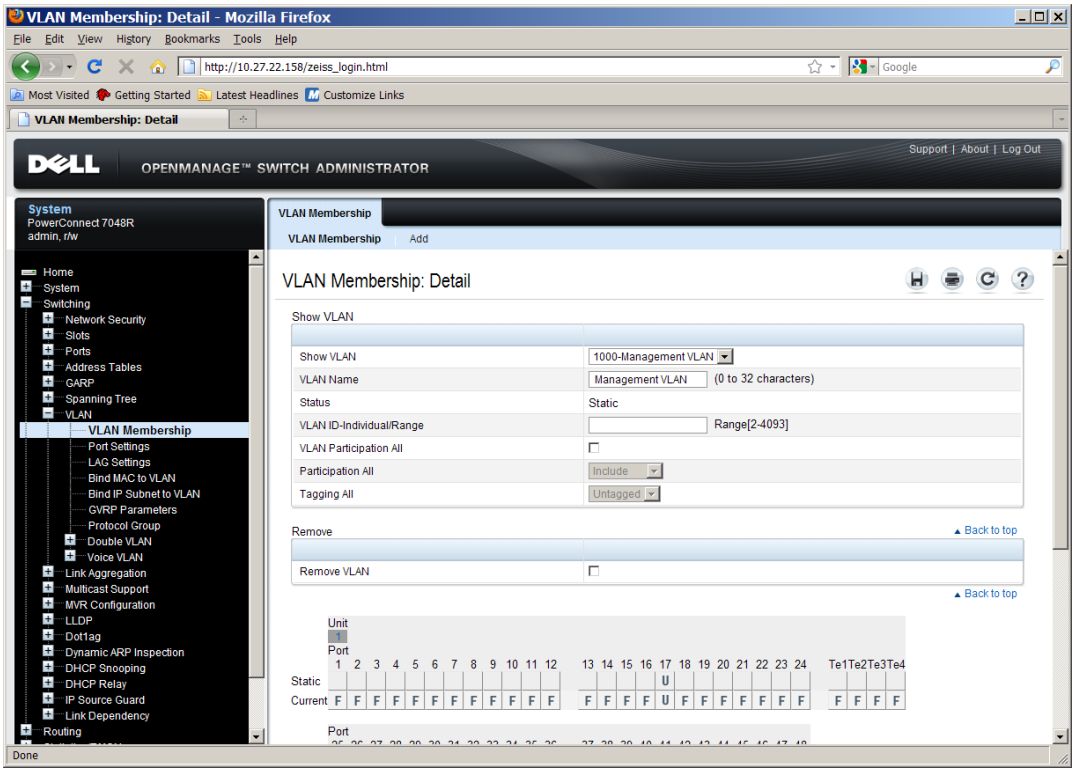
2. From the **Add VLAN** page, enter **1000** in the VLAN ID field and click **Apply**.



3. From the Ports menu on the **Switching > VLAN > Port Settings** page, select port **Gi1/0/17**.
4. Configure port gi1/0/17 in General mode with a PVID of **1000** and click **Apply Changes**.

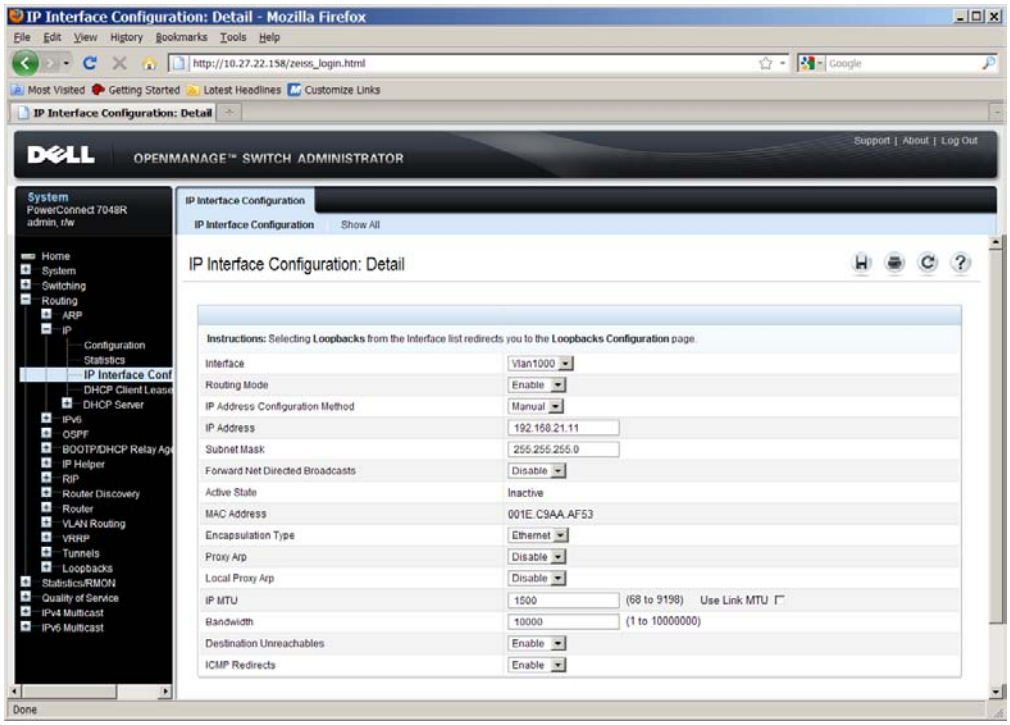


5. From the **Show VLAN** menu on the **Switching > VLAN > VLAN Membership** page, select **1000**.
6. Click the **Static** box for port 17 so that the letter **U** (untagged) appears in the box.
7. Click **Apply**.

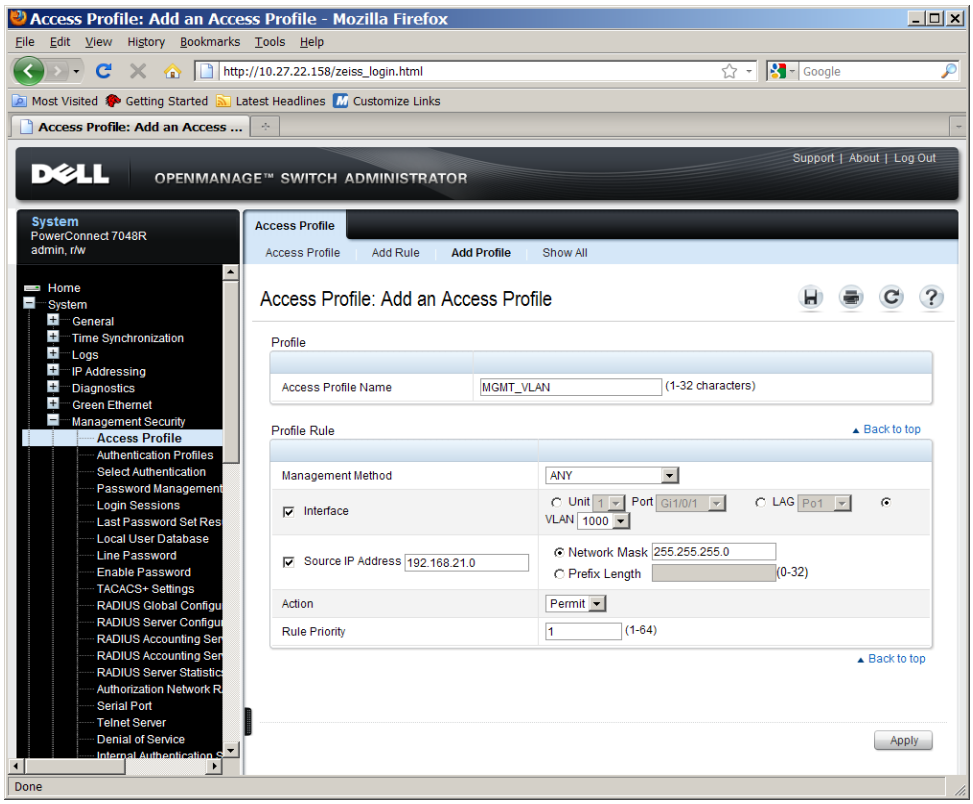


8. Navigate to the **Routing > IP > IP Interface Configuration** page.

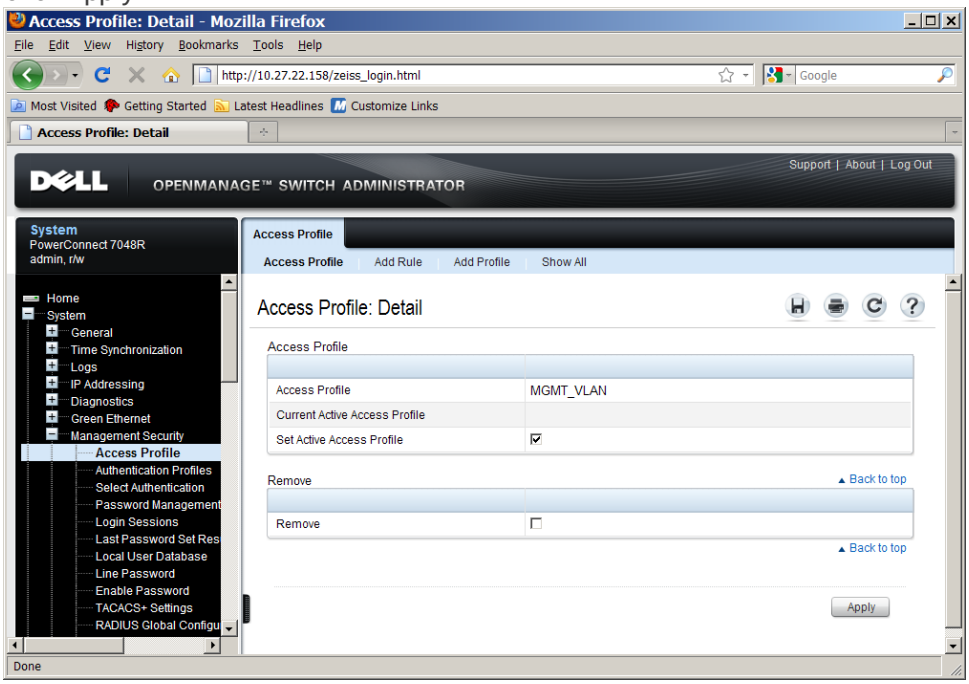
9. In the **Interface** field, select **VLAN 1000**. In the **Routing Mode** field, select **Enable**. In the **IP Address Configuration Method** field, select **Manual.** In the **IP Address** field, enter **192.168.21.11.** In the **Subnet Mask** field, enter **255.255.255.0.** Click **Apply.**



10. Navigate to the **System > Management Security > Access Profile** page. Choose the **Add Profile** tab. Enter **MGMT_VLAN** in the **Access Profile Name** field. Select the **Interface** check box and the **VLAN** radio button. Choose **1000** in the **VLAN** drop down. Enable the **Source IP Address** check box and enter 192**.168.21.0** in the **Source Ip Address** field. Enter **255.255.255.0** in the **Network Mask** field. Ensure that the **Action** field is set to **Permit** and set the **Rule Priority** field to **1**.



11. Click **Apply**.
12. Navigate to the **System > Management Security > Access Profile** page. Choose the **Access Profile** tab. Enable the **Set Active Access Profile** check box in the **MGMT_VLAN** Access Profile and click Apply.



Adding a Stack Member with Minimal Interruption

When adding a new member to a stack, make sure that only the stack cables and no host-facing network cables are connected before powering up the new unit. Make sure the end-user links are not connected to any ports on the unit being added and the new member is powered off. This is important because if STP is enabled and any links are UP, then Spanning Tree Protocol (STP) re-convergence will take place as soon as the link is detected.

After the stack cables on the new member are connected to the stack, connect the switch power. Do not connect another switch to the stack until the existing members are powered up. Also, do not connect two functional, powered-up stacks together. If two functional, powered-up stacks are connected together or connect a powered-up new member to the stack, then master re-election takes place, which causes the stack that no longer has a master to reboot. See “Merging Two Operational Stacks” for more details.

An “unassigned unit” is a switch that is preconfigured to be a member of a stack, but has not been physically connected to the stack. If there are any unassigned units already configured on the stack, remove them prior to adding a new unit to stack. This is important because when there is any preconfigured unit and the master holds some configuration for that unit, as soon as the new unit is detected, the configuration is applied, which might trigger the re-convergence or startup of many other protocols. However, it is possible to intentionally pre-configure a unit. The preconfigured/unassigned units can be viewed by using the show switch CLI command shown below.

The following example shows how to view the units in the stack and remove an unassigned unit:

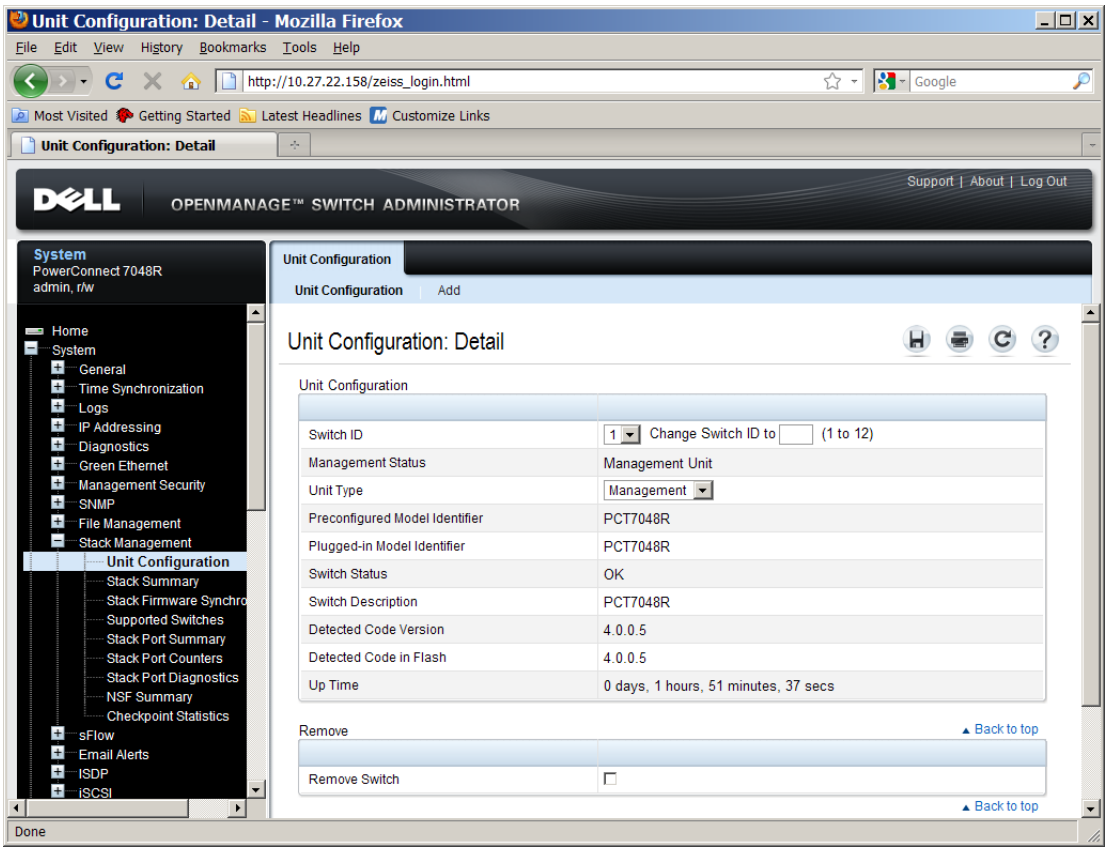
```
console#show switch

    Management Standby   Preconfig     Plugged-in    Switch        Code
SW  Status     Status    Model ID      Model ID      Status        Version
--- ---------- --------- ------------- ------------- ------------- ----------
1   Mgmt Sw              PCT7048P      PCT7028P      OK            4.0.0.6
2   Unassigned           PCT7028P                    Not Present   0.0.0.0

console#configure
console(config)#stack
console(config-stack)#no member 2
console(config-stack)#exit
console(config)#exit
```

Removing a Stack Member with Minimal Interruption

NOTE: This migration process can be disruptive, so schedule some outage time. To minimize any disruption, plan on having one or more test scenarios prepared in order to verify that the migrated ports are operating properly.

1. Migrate end-user ports to appropriately configured ports on a different stack member. Verify that the migrated end-user services are operational. If an end user is not operational, it is possible to recover quickly by connecting the end user back to the original port.
2. Migrate any trunk ports to an appropriately configured port on a different stack member. Verify the operation of the new trunk port. This may require several steps:
    a. Remove spanning tree configuration from the ports on the stack member to be removed.
    b. Remove LAG configuration from the ports on the stack member to be removed.
    c. Reroute any statically routed traffic going through the stack member.
3. Remove the member from the stack, re-cable around the member, checking that the cabling is positively connected and does not show stacking errors. Power off the removed member.

To remove a switch from the stack by using the Web interface, navigate to the **System > Stack Management > Unit Configuration** page. Select the switch to remove from the Switch ID drop down, and then select the **Remove Switch** option. Then, click **Apply**.



Merging Two Operational Stacks

The recommended procedure for merging two operational stacks is as follows:

Caution! This procedure is disruptive and requires a network outage.

1. Always power off all units in one stack before connecting to another stack.
2. Add the units as a group by unplugging one stacking cable in the operational stack and physically connecting all unpowered units.
3. Completely cable the stacking connections, making sure the redundant link is also in place.

Two operational stacks can also be merged by reconnecting stack cables without powering down units in either stack. Connecting a powered-up standalone unit to an existing stack leads to same behavior as when merging two operational stacks. In such cases, manager election is performed and the manager unit with the higher MAC address is chosen as manager. The stack that no longer has a stack manager resets itself and all of its member units. After the reset, all the stack members in the reset stack join the stack that has the manager to form a single stack. The stack that has the manager remains functional through the merge process.

Synchronizing the Running Configuration between the Master and Standby Units

The master unit copies its running configuration to the standby unit whenever the configuration changes (subject to some minimal time delays introduced in the copy mechanism to increase the efficiency of the transfer process). This enables the standby unit to take over the stack operation with minimal interruption if the master unit becomes unavailable.

The running-config synchronization also occurs:

- When the administrator saves the running configuration to the startup configuration on the master unit.
- When the standby unit changes.

Master Failover

If the current master unit fails, the standby unit becomes the master unit. If no switch is preconfigured as the standby unit, the firmware automatically selects a standby unit from the existing stack units.

When the failed master resumes normal operation, it joins the stack as a member (not a master) if the new master unit has already been elected.

The stack supports nonstop forwarding and graceful restart during a master failover. See “Nonstop Forwarding” on page 22.

Effect of Master Failover on PoE Devices

Stack members that provide power-over-Ethernet (PoE) to connected devices continue to provide power in the event of a master failover. The PoE application uses a separate controller on each unit, and these controllers are not initialized upon a warm restart.

If PoE is enabled on a switch port shortly before a master failover and the hardware status change had not yet been communicated to the backup unit, the switch will continue to provide PoE power even though PoE is not enabled for the port in the configuration on the new master. The user can correct this configuration mismatch by re-enabling PoE on the port.

Stack Member Failover

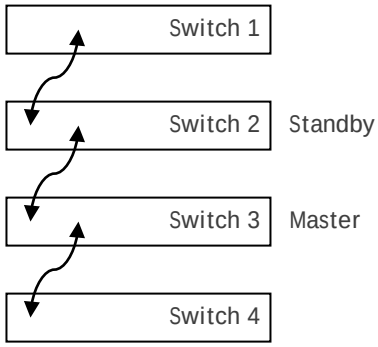
When a unit in the stack fails, the master unit removes the failed unit from the stack. No changes or configuration settings are applied to the other stack members; however, dynamic protocols will try to re-converge as the topology could change because of the failed unit. When there are no connected ports on the failed unit, the stack will be intact without any changes.

Failover Scenarios

This section describes examples of what happens when a stack member or the master unit fails.

Scenario 1

In this example, the stack has four members that are connected through a daisy-chain.



When all four units are up and running, the `show switch` CLI command gives the following output:

```
console#show switch

       Managemen  Standby       Preconfig Plugged-in  Switch       Code
Switch Status     Status        Model ID  Model ID    Status       Version
------ ---------- ------------- --------- --------    ------------ --------
1      Stack Mbr  Standby       PCT7048   PCT7048     OK           4.0.0.6
2      Stack Mbr                PCT7048   PCT7048     OK           4.0.0.6
3      Mgmt Sw                  PCT7048   PCT7048     OK           4.0.0.6
4      Stack Mbr                PCT7048   PCT7048     OK           4.0.0.6
```

At this point, if Unit 2 is powered off or rebooted due to an unexpected failure, `show switch` gives the following output:

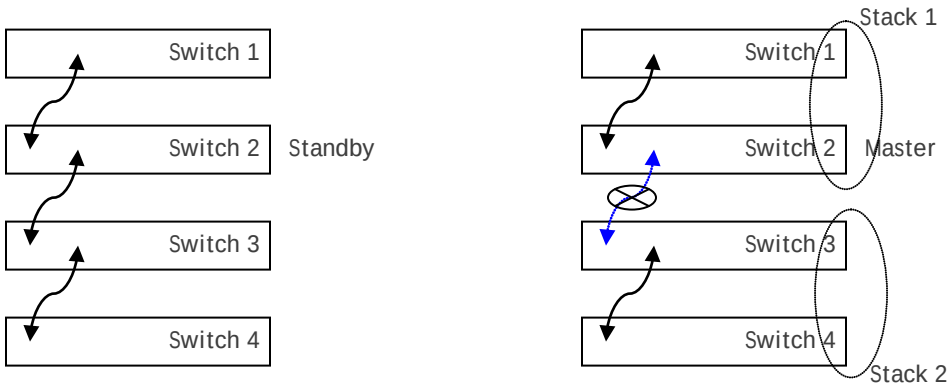
```
console#show switch

       Management  Standby       Preconfig Plugged-in  Switch       Code
Switch Status      Status        Model ID  Model ID    Status       Version
------ ----------- ------------  --------- ----------- ------------ --------
1      Stack Mbr   Standby       PCT7048   PCT7048     OK           4.0.0.6
2      Unassigned                PCT7048               Not Present  0.0.0.0
3      Mgmt Switch               PCT7048   PCT7048     OK           4.0.0.6
4      Stack Mbr                 PCT7048   PCT7048     OK           4.0.0.6
```

When the failed unit resumes normal operation, the previous configuration that exists for that unit is reapplied by the master unit.

Scenario 2

Consider the same example with a four-unit stack connected in daisy-chain fashion.

Figure 4. Stack Split



If the link between Switch 2 and Switch 3 is removed, the stack is split into two different stacks with Switches 1 and 2 in one stack and Switches 3 and 4 in another.

The master unit for each stack is determined by the following criteria:

- Switch 3 was configured as master prior to the split, so it will continue to be the master unit for Stack 2 (units 3 and 4).
- If Switch 2 is configured as the standby, it becomes the master for Stack 1 (units 1 and 2).
- If none of the units in Stack 1 are configured as master or standby, then Stack 1 will have no master after the split. The election process will start on these two units, and either Switch 1 or Switch 2 will come up as a master based on which switch has the highest base MAC address.

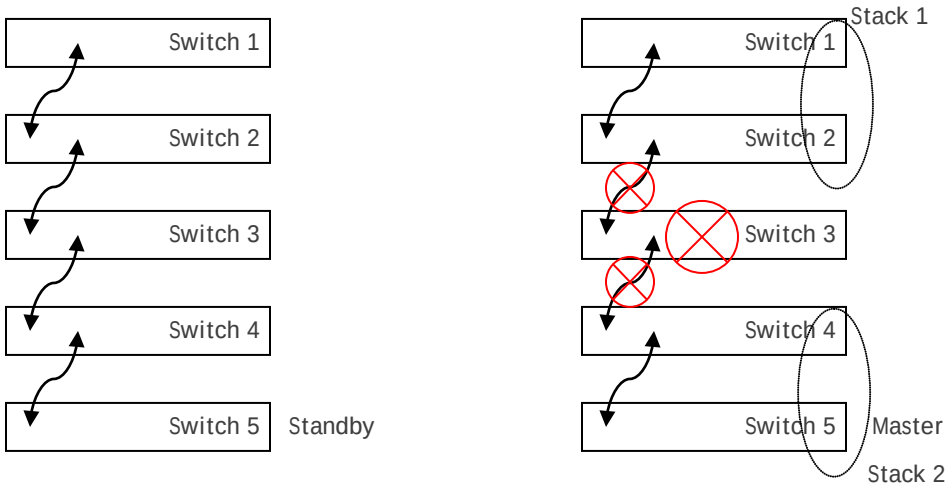
In this specific example, there will not be any change in the configuration in Stack 2. Stack 1 will come up with the previously saved configuration on the new master elected.

Note that stack splits are less likely when the switches are cabled in a ring. For example, if Switch 4 was connected to Switch 1 and the link between Switch 2 and Switch 3 was removed, all switches would remain members of the single stack with Switch 3 continuing as the master.

Scenario 3

The following example contains a similar condition with a master unit failover and consequent stack split with five units in the stack, as Figure 5 shows.

Figure 5. Stack Split with Manager Failure



In this example, there is no link back to Switch 5 from Switch 1. In this case, if the manager of the stack goes down (failed/rebooted), the stack is split into two different stacks with units 1 and 2 in one stack and units 4 and 5 in another. With this condition, none of the stacks will have a working master within the stack, so both of the stacks will elect new masters for each stack through the process based on which unit has the highest base MAC address.

Nonstop Forwarding

A switch can be described in terms of three semi-independent functions called the forwarding plane, the control plane, and the management plane:

- **Forwarding Plane** — The set of hardware components that forward data packets without intervention from a control CPU, sometimes called the Data Plane. The forwarding plane is implemented in hardware.
- **Control Plane** — The firmware layer that manages system and hardware configuration and runs the network control protocols in order to set system configuration and state. The control plane determines how the forwarding plane should forward packets, deciding which data packets are allowed to be forwarded and where they should go. The control plane is implemented in application firmware running on the management unit.
- **Management Plane** — A set of interfaces that enable the network administrator to configure the networking device. The management plane is implemented in application firmware running on the management unit.

Nonstop forwarding (NSF) allows the forwarding plane of stacked units to continue to forward packets while the control and management planes restart as a result of a power failure, hardware failure, or firmware fault on the management unit. A nonstop forwarding failover can also be manually initiated using the `initiate failover` command. Traffic flows that enter and exit the stack through physical ports on a unit other than the management continue with virtually no interruption when the management unit fails. To prepare the standby management unit in case of a failover, applications on the management unit periodically checkpoint state information to the standby unit. Changes to the running configuration are automatically copied to the standby unit. The MAC address for each switch in the stack stays the same across a nonstop forwarding failover, so that neighbors do not have to relearn them.

For NSF to be effective, adjacent networking devices must not reroute traffic around the restarting device. Three techniques are used to prevent traffic from being rerouted:

1. A protocol may distribute a part of its control plane to stack units so that the protocol can give the appearance that it is still functional during the restart. When NSF is enabled on a switch, various protocols and configurations such as Spanning Tree and Link Access groups automatically use this technique.
2. A protocol may enlist the cooperation of its neighbors through a technique known as graceful restart. Graceful restart functionality can be enabled for the OSPF and OSPFv3 protocols so that the stack can continue to forward packets using the same IPv4 and IPv6 routes while the standby unit takes over management responsibility. For example, when OSPF executes a graceful restart, it informs its neighbors that the OSPF control plane is restarting, but that it will be back shortly. Helpful neighbors continue to advertise to the rest of the network that they have full adjacencies with the restarting router, avoiding announcement of a topology change and the network activity that would result (i.e., flooding of LSAs, SPF runs). Helpful neighbors continue to forward packets through the restarting router. The restarting router relearns the network topology from its helpful neighbors.
3. A protocol may simply restart after the failover if neighbors react slowly enough that they will not normally detect the outage. The IP multicast routing protocols are a good example of this behavior in that the PIM stack restarts before its neighbors detect its absence and drop any adjacencies.

To take full advantage of nonstop forwarding, layer 2 connections to neighbors should be via port channels that span two or more stack units, and layer 3 routes should be equal-cost multi-path (ECMP) routes. ECMP provides load balancing, using multiple next hops to a single destination via physical ports on two or more units. The hardware can quickly move traffic flows from port channel members or ECMP paths on a failed unit to a surviving unit.

Initiating a Warm Failover of the Manager Unit

The administrator can use the `initiate failover` command to initiate a "warm" restart. This command reloads the management unit, triggering the standby unit to take over. As the standby management unit takes over, the system continues to forward end-user traffic. The end-user data streams switched over the failing switch may lose a few packets during the failure, but they do not lose their IP sessions, such as VoIP calls.

If no standby unit is available when the `initiate failover` command is issued, the command fails with an error message stating that no standby unit exists. If the standby unit is not ready for a warm restart, the command fails with a similar error message.Administrators should be aware that use of the `movemanagement` command is not recommended in operational networks as network traffic interferes with stack convergence. The `movemanagement` command performs a full synchronization and then triggers a cold restart, even if the target unit is the backup unit, and is therefore less preferred than the `initiate failover` command.

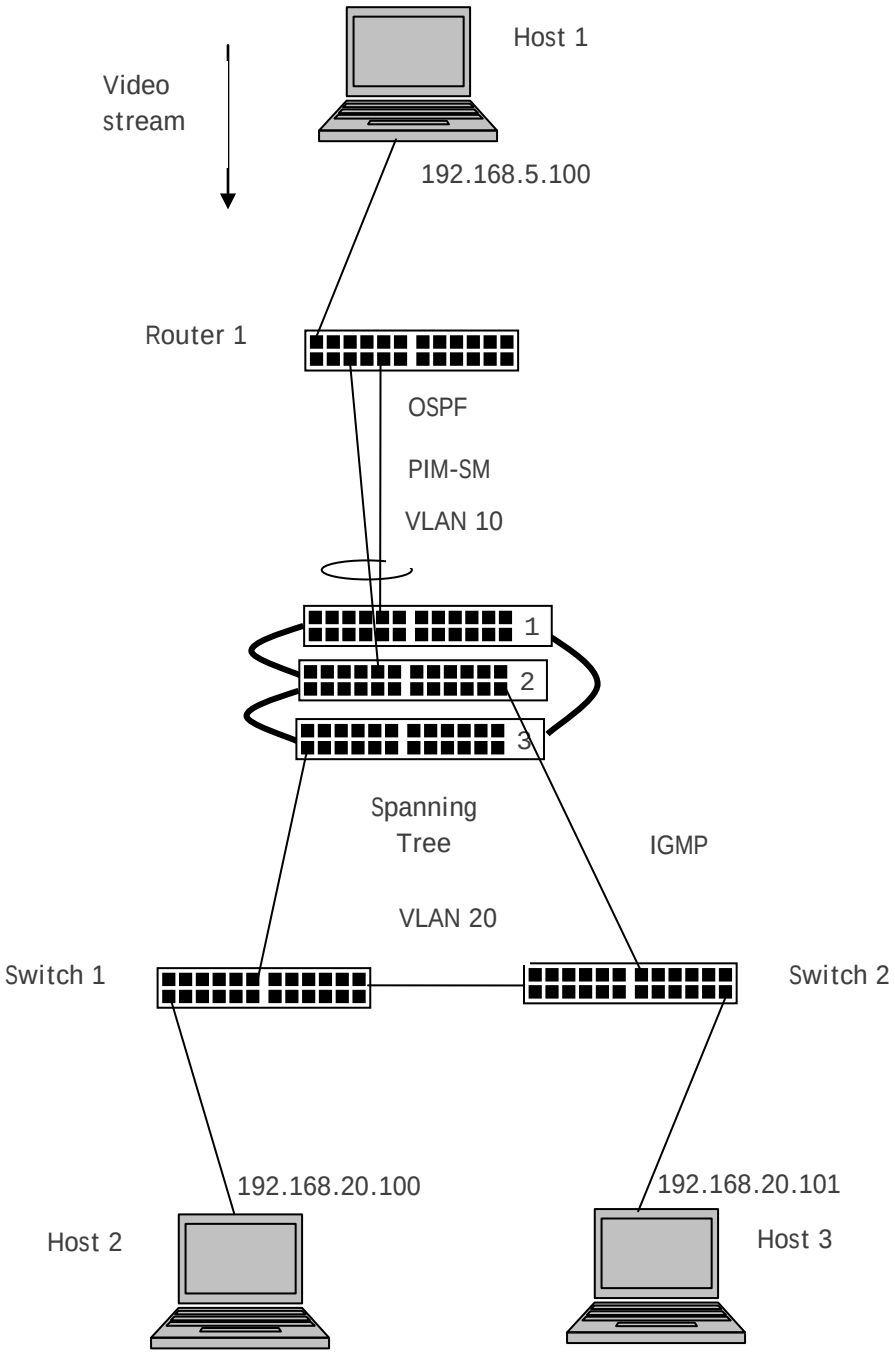
Nonstop Forwarding Scenario

Figure 6 depicts the following network setup:

- Host 1 is a video stream server. Hosts 2 and 3 are video stream receivers.
- The stack has OSPF and PIM-SM adjacencies with Router 1 over a LAG with member ports in unit 1 and unit 2.
- Spanning tree runs on the links in the L2 network connecting the stack, Switch 1, and Switch 2.
- The stack runs IGMP on the links toward Host 2 and Host 3.
- VLAN 20 is a routing VLAN with IP address 192.168.20.2 on the stack.
- The network is configured to select Switch 2 as the root bridge. The stack selects its direct link to Switch 2 as its root port. The stack puts its link to Switch 1 in the Discarding state.
- Router 1 is a Cisco router that serves as a static rendezvous point (RP) in the PIM-SM network.

The CLI for configuring this scenario follows the illustration, along with a description of how the stack responds during a failover with NSF enabled and disabled.

Figure 6. Nonstop Forwarding Example



NSF Scenario Configuration via CLI

Router 1:

```
configure terminal
hostname Router_1
ip multicast
ip pim rp-address 192.168.10.1 239.0.1.1 255.255.255.0
ip pim sparse

vlan database
 vlan 10
exit

interface vlan 10
 ip address 192.168.10.1 255.255.255.0
 ip pim
exit

router ospf
 router-id 1.1.1.1
 network 192.168.0.0 0.0.255.255 area 0
 enable
exit

interface Port-channel1
 description “LAG to switch stack – gi1/0/10, gi1/0/20”
 switchport mode trunk
 switchport trunk native vlan 10
 switchport trunk allowed vlan remove 1
exit

interface GigabitEthernet 1/0/1
 description “Interface to multicast source”
 no spanning-tree
 switchport access vlan 10
exit

interface GigabitEthernet 1/0/10
 no spanning-tree auto-portfast
 channel-group 1 mode auto
exit

interface GigabitEthernet 1/0/20
 no spanning-tree auto-portfast
 channel-group 1 mode auto
exit


----------------------------------------
```

Stack:

```
configure terminal
vlan database
 vlan  10,20
exit
vlan 10
 name “Routed interface to PIM RP via port-channel 1”
exit
vlan 20
 name “Routed interface to downstream switched network”
exit

hostname "nsf-stack"
spanning-tree mode rst
ip igmp
ip pim sparse
ip multicast
ip pim rp-address 192.168.10.1 239.0.1.1 255.255.255.0

interface Port-channel1
 description “LAG to PIM router – VLAN 10 – gi1/0/10, gi2/0/10”
 switchport mode trunk
 switchport trunk native vlan 10
 switchport trunk allowed vlan remove 1
exit

interface vlan 10
 ip address  192.168.10.2  255.255.255.0
 ip igmp
 ip pim
exit

interface vlan 20
 ip address  192.168.20.2  255.255.255.0
 ip igmp
 ip pim
exit

router ospf
 router-id 2.2.2.2
 network 192.168.0.0 0.0.255.255 area 0
 nsf
exit

interface gi1/0/10
 no spanning-tree auto-portfast
 channel-group 1 mode auto
```

```
exit

interface gi2/0/10
 no spanning-tree auto-portfast
 channel-group 1 mode auto
exit

interface gi2/0/20
 switchport mode general
 switchport general pvid 20
 switchport general acceptable-frame-type tagged-only
 switchport general allowed vlan add 20 tagged
 switchport general allowed vlan remove 1
exit


interface ethernet gi3/0/20
 switchport mode general
 switchport general pvid 20
 switchport general acceptable-frame-type tagged-only
 switchport general allowed vlan add 20 tagged
 switchport general allowed vlan remove 1
exit

exit
```

Switch 1

```
configure terminal
hostname Switch-1
vlan database
 vlan 20
exit

spanning-tree mode rst
spanning-tree priority 16384
interface gi1/0/20
 switchport mode general
 switchport general pvid 20
 switchport general acceptable-frame-type tagged-only
 switchport general allowed vlan add 20 tagged
 switchport general allowed vlan remove 1
exit

interface gi1/0/1
 switchport mode access
 switchport access vlan 20
exit
```

```
interface gi1/0/5
 switchport mode general
 switchport general pvid 20
 switchport general acceptable-frame-type tagged-only
 switchport general allowed vlan add 20 tagged
 switchport general allowed vlan remove 1
exit
exit
```

Switch 2

```
configure terminal
hostname Switch-2
vlan database
 vlan 20
exit

spanning-tree mode rst
spanning-tree priority 12288
interface gi1/0/20
 switchport mode general
 switchport general pvid 20
 switchport general acceptable-frame-type tagged-only
 switchport general allowed vlan add 20 tagged
 switchport general allowed vlan remove 1
exit

interface gi1/0/1
  switchport mode access
  switchport access vlan 20
exit

interface gi1/0/5
 switchport mode general
 switchport general pvid 20
 switchport general acceptable-frame-type tagged-only
 switchport general allowed vlan add 20 tagged
 switchport general allowed vlan remove 1
exit
exit
```

NSF Example 1

Assume Unit 1 is the stack manager and Unit 2 is the standby switch. The following scenarios illustrate how the network recovers when Unit 1 is powered down.

- **With NSF disabled:** Unit 2 takes over as manager with a cold restart and clears the hardware tables. The video stream stops on Hosts 2 and 3 for at least 30 seconds as the OSPF adjacency is rebuilt, multicast routes are relearned, and spanning tree reconverges.
- **With NSF enabled:** Unit 2 takes over as manager with a warm restart. OSPF graceful restart keeps the adjacency with Router 1 up and Router 1 continues to forward to the stack. There is no perceivable outage of the video stream through the stack.

NSF Example 2:

Assume Unit 2 is the stack manager and Unit 3 is the standby switch, and both ports are active on the LAG. In this scenario, NSF is enabled and the `initiate failover` command is used.

- Because the unit with the root port goes down, the stack stops forwarding until the spanning tree control plane comes back and places the link on Unit 3 in the forwarding state (about 3 seconds).

NSF Re-convergence Timing

As mentioned in the previous section, NSF protects against failures by check-pointing information to a standby unit. In an NSF-protected stack, the worst-case scenario is when the master unit fails. The following statistics show representative re-convergence times of an NSF-enabled standby switch as measured from the detection of failure. With NSF enabled, data plane forwarding continues non-stop on the non-failed switches while the standby unit converges the control plane protocols using the check-pointed information. The re-convergence times given below include re-establishing communication with the upper layer protocol peers, synchronization of shared information with the peer, and re-establishment of any data plane forwarding paths around the failed master unit.

Medium Configuration

- 8 switches stacked
- 100 VLANs, all ports are members of all VLANs
- 4/4 static/dynamic lags with 8 members each
- 3 MSTP instance with VLANs
- 30 ACLs applied on 30 interfaces
- 10 Diffserv service interfaces with policies on 10 interfaces
- 6 VLAN routing interfaces
- 128 L2 Multicast group entries
- 512 ARP entries
- 128 Unicast routes

Parameter	Timing
L2 loss duration (non-failed or rerouted stack member)	0 msec
L2 loss duration (failed/rerouted stack member)	10 msec
L3 loss duration (failed/rerouted stack member)	12 msec
IPMC loss duration (failed/rerouted stack member)	12 msec
L2 convergence time	25.70 sec
L3 convergence time	25.410 sec
IPMC convergence time	25.420 sec
Total convergence time	25.420 sec

Small Configuration

- 6 switches stacked
- 4 VLANs, all ports are members of all VLANs
- 2/2 static/dynamic lags with 8 members each
- 1 MSTP instance with VLANs
- 15 ACLs applied on 15 interfaces
- 5 Diffserv service interfaces with policies on 5 interfaces
- 2 VLAN routing interfaces
- 16 L2 Multicast group entries
- 256 ARP entries
- 16 Unicast routes

Parameter	Timing
L2 loss duration (non-failed or rerouted stack member)	0 msec
L2 loss duration (failed/rerouted stack member)	10 msec
L3 loss duration (failed/rerouted stack member)	12 msec
IPMC loss duration (failed/rerouted stack member)	12 msec
L2 convergence time	14.770 sec
L3 convergence time	14.800 sec
IPMC convergence time	17.540 sec
Total convergence time	17.540 sec

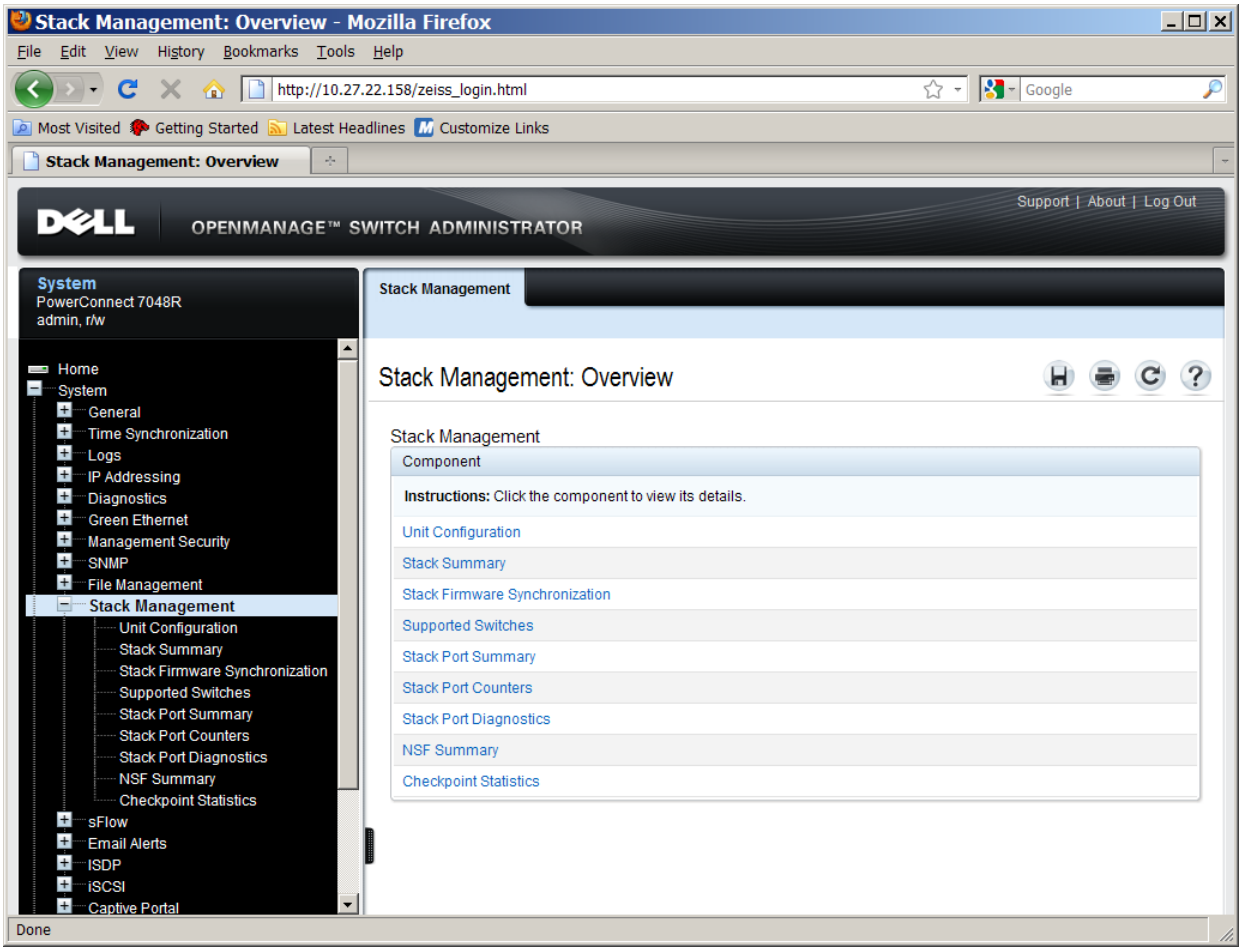
Stacking CLI Commands

The following stacking commands are available from the CLI. See the CLI Reference Guide for more detailed information about the commands.

```
boot auto-copy-sw
clear checkpoint statistics
initiate failover
member
movemanagement
nsf
set description
show checkpoint statistics
show nsf
show stack-standby
show switch stack-ports
show switch stack-ports counters
show switch stack-ports diag
show supported switchtype
show switch
stack-port
switch renumber
```

Stacking Web Interface

The stacking configuration and monitoring Web pages are accessed from the **System > Stack Management** menu. The following image shows a list of the Web pages that are available for the stacking feature.



Summary

This paper has described the theory and operation of stacking on the Dell PowerConnect 7000 series of switches. Delivering significant rack density, the PowerConnect 7000 gives network administrators the flexibility to maximize server and workstation connectivity in a 1U form factor. Stacking provides the ultimate in ease of use and manageability with automatic firmware version synchronization and a single management interface for up to 12 switches. Stacked switches can also be managed via Dell OpenManage™ IT Assistant and Dell OpenManage Network Manager, as well as third-party SNMP-based management console applications. Stacks with link- and switch-level redundancy and quick failover times are capable of meeting the most demanding high availability requirements of today's global enterprises and data centers. Most importantly, stacking provides the ability to grow your switching capacity as your business grows. In summary, the value of stacking helps reduce your total cost of ownership (TCO), increase your business agility, and improve your network resiliency.