High Availability

This chapter describes how to identify and resolve problems related to high availability.

Information About High Availability

The purpose of high availability (HA) is to limit the impact of failures—both hardware and software—within a system. The Cisco NX-OS operating system is designed for high availability at the network, system, and service levels.

The following Cisco NX-OS features minimize or prevent traffic disruption in the event of a failure:

- Redundancy—Redundancy at every aspect of the software architecture.
- Isolation of processes—Isolation between software components to prevent a failure within one process that is disrupting other processes.
- Restartability—Most system functions and services are isolated so that they can be restarted independently after a failure while other services continue to run. In addition, most system services can perform stateful restarts, which allow the service to resume operations transparently to other services.
- Supervisor stateful switchover— Active/standby dual supervisor configuration. The state and configuration remain constantly synchronized between two Virtual Supervisor Modules (VSMs) to provide a seamless and statefu1 switchover in the event of a VSM failure.

The Cisco Nexus 1000V system is made up of the following:

- Virtual Ethernet Modules (VEMs) that run within virtualization servers. The VEMs are represented as modules within the VSM.
- A remote management component, for example, the Microsoft System Center Virtual Machine Manager (SCVMM).
- One or two VSMs that run within Virtual Machines (VMs).

Problems with High Availability

Symptom	Possible Causes	Solution
The active VSM does not see the standby VSM.	Roles are not configured properly. • Check the role of the two VSMs by entering the **show system redundancy status** command.	1. Confirm that the roles are the primary and secondary role, respectively. 2. If needed, enter the **system redundancy role** command to correct the situation. 3. Save the configuration if roles are changed.
	Network connectivity problems. • Check the control and management VLAN connectivity between the VSM at the upstream and virtual switches.	If network problems exist, do the following: 1. From the Microsoft SCVMM UI, shut down the VSM, which should be in standby mode. 2. From the Microsoft SCVMM UI client, bring up the standby VSM after network connectivity is restored.
The active VSM does not complete synchronization with the standby VSM.	Version mismatch between VSMs. • Check that the primary and secondary VSM are using the same image version by entering the **show version** of the command.	If the active and standby VSM software versions differ, reinstall the secondary VSM with the same version used in the primary.
	Fatal errors during gsync process. • Check the gsyncctrl log by entering the **show system internal log sysmgr gsyncctrl** command and look for fatal errors.	Reload the standby VSM by entering the **reload module** *module-number* command where *module-number* is the module number for the standby VSM.

Symptom	Possible Causes	Solution
The standby VSM reboots periodically.	The VSM has connectivity only through the management interface. • When a VSM is able to communicate through the management interface, but not through the control interface, the active VSM detects the situation and resets the standby VSM to prevent the two VSMs from being in HA mode and out of sync. • Check the output of the **show system internal redundancy info** command and verify if the *degraded_mode* flag is set to true.	Check the control VLAN connectivity between the primary and secondary VSMs.
	VSMs have different versions. Enter the **debug system internal sysmgr all** command and look for the active_verctrl entry that indicates a version mismatch, as the following output shows: `2009 May 5 08:34:15.721920 sysmgr: active_verctrl: Stdby running diff version- force download the standby sup.`	Isolate the standby VSM and boot it. Enter the **show version** command to check the software version in both VSMs. Install the image matching the active VSM on the standby.

Symptom	Possible Causes	Solution
Both VSMs are in active mode.	Network connectivity problems. • Check for control and management VLAN connectivity between the VSM at the upstream and virtual switches. • When the VSM cannot communicate through any of these two interfaces, they will both try to become active.	If network problems exist, do the following: 1. From the Microsoft SCVMM UI client, shut down the VSM, which should be in standby mode. 2. From the Microsoft SCVMM UI client, bring up the standby VSM after network connectivity is restored.
	Different domain IDs in the two VSMs. Check the *domain* value by entering the **show system internal redundancy info** command.	If needed, update the domain ID and save it to the startup configuration. • Upgrading the domain ID in a dual VSM system must be done following this procedure: – Isolate the VSM with the incorrect domain ID so that it cannot communicate with the other VSM. – Change the domain ID in the isolated VSM, save configuration, and power off the VSM. – Reconnect the isolated VSM and power it on.

System-Level High Availability

The Cisco Nexus 1000V supports redundant VSM VMs—a primary and a secondary—that run as an HA pair. Dual VSMs operate in an active/standby capacity in which only one of the VSMs is active at any given time, while the other acts as a standby backup. The state and configuration remain constantly synchronized between the two VSMs to provide a stateful switchover if the active VSM fails.

Single or Dual Supervisors

The Cisco Nexus 1000V system is made up of the following:

- VEMs that run within virtualization servers (these VEMs are represented as modules within the VSM)
- A remote management component, such as the Microsoft SCVMM.
- One or two VSMs that run within VMs.

Single VSM Operation	Dual VSM Operation
• Stateless—Service restarts from the startup configuration • Stateful—Service resumes from previous state.	• One active VSM and one standby VSM. • The active VSM runs all the system applications and controls the system. • On the standby VSM, the applications are started and initialized in standby mode. They are also synchronized and kept up to date with the active VSM in order to maintain the runtime context of "ready to run." • On a switchover, the standby VSM takes over for the active VSM.

Network-Level High Availability

The Cisco Nexus 1000V HA at the network level includes port channels and the Link Aggregation Control Protocol (LACP). A port channel bundles physical links into a channel group to create a single logical link that provides the aggregate bandwidth of up to eight physical links. If a member port within a port channel fails, the traffic that was previously carried over the failed link switches to the remaining member ports within the port channel.

Additionally, the LACP allows you to configure up to 16 interfaces into a port channel. A maximum of eight interfaces can be active, and a maximum of eight interfaces can be placed in a standby state.

For additional information about port channels and the LACP, see the *Cisco Nexus 1000V for Microsoft Hyper-V Layer 2 Switching Configuration Guide*.

Failover Clusters and the Microsoft SCVMM

Failover clustering is a hostside feature that provides high availability and scalability to multiple server workloads. In order for a Cisco Nexus 1000V switch to be considered a high availability device, the switch must meet the following criteria:

- The VM must be set to **High Availability** > **True** to be considered part of a failover cluster. That is, the VM can be moved automatically by the cluster in the event of a host failure.
- The high availability VM should be stored in one of the following types of Internet Protocol (IP) based storage facilities to accommodate live migration for a failover cluster:
  - Shared SMB storage
  - Clustered shared volumes (iSCSI, and so on)

When clusters are managed by the Microsoft SCVMM, certain criteria must be met for the Microsoft SCVMM to manage the VM as part of a failover cluster. That is, the logical switch that is part of the hosts of the failover clusters should be configured for high availability.

High Availability Logical Switch Criteria and Behavior

- A logical switch is considered to be highly available when it carries the same uplink networks on all the nodes of the cluster.
- If certain adapters carry the same uplink in each logical switch across all nodes and other uplinks do not then the adapters that carry the same uplink networks become high availability.

- A VM that is not configured for high availability can be connected to any switch in the failover cluster (logical or standard switch).
- A high availability VM can only be connected to uplinks that are high availability and are part of a logical switch.

Selecting Storage During VM Deployment on Failover Clusters from the Microsoft SCVMM

The failover cluster managed by the Microsoft SCVMM has more than one associated storage device. By default, the Microsoft SCVMM chooses the storage based on the deployment algorithm of the Microsoft SCVMM, which might not be what you want.

Step 1 Launch the Microsoft SCVMM UI.

Step 2 In the **Migrate VM Wizard** screen, change the storage of the VM and the VM hard disk to the appropriate storage.

Step 3 Pin the selection to the Microsoft SCVMM UI.

Live Migration Fails Due to Network Bandwidth

When a workload VM is carrying high traffic, VM live migration might not be allowed by the Microsoft SCVMM. The Microsoft SCVMM performs checks during live mgration and decides the feasibility of moving the VM based on many factors, one of which is VM port traffic. From the perspective of the Microsoft SCVMM, when a VM is transmitting or receiving large amounts of traffic, it is not feasible to move the VM because it might result in a loss of bandwidth.

Step 1 Launch the Microsoft SCVMM UI.

Step 2 In a Microsoft SCVMM PowerShell window, enter **Move-SCVirtualMachine**.

Cluster IP Resource Fails to Come Up

Cluster validation is an important tool used by large deployments to validate cluster configurations. When a virtual switch is deployed on the management NIC of the host with a static IP address, and the failover cluster already exists, the cluster IP resource might fail to come up. When this problem occurs, although the cluster IP address and DNS are reachable by conventional means (ping), the cluster validation tool fails.

Note This problem is a known issue with the Microsoft SCVMM and is seen only with static IP addresses, not when the host management IP address is distributed over DHCP.

There is no known workaround for this issue. We recommended that you create clusters after you deploy the Cisco Nexus1000V on the management IP address.

High Availability Troubleshooting Commands

You can use the commands in this section to troubleshoot problems related to high availability.

To list process logs and cores, enter these commands:

- **show cores**

```
switch# show cores
Module  Instance  Process-name     PID       Date(Year-Month-Day Time)
------  --------  ---------------  --------  -------------------------
1       1        private-vlan      3207    Apr 28 13:29
```

- **show processes log [pid** *pid*]

```
switch# show processes log
Process          PID     Normal-exit  Stack  Core   Log-create-time
---------------  ------  -----------  -----  -----  ---------------
private-vlan     3207              N      Y      N  Tue Apr 28 13:29:48 2009

switch# show processes log pid 3207
==================================================
Service: private-vlan
Description: Private VLAN

Started at Wed Apr 22 18:41:25 2009 (235489 us)
Stopped at Tue Apr 28 13:29:48 2009 (309243 us)
Uptime: 5 days 18 hours 48 minutes 23 seconds

Start type: SRV_OPTION_RESTART_STATELESS (23)
Death reason: SYSMGR_DEATH_REASON_FAILURE_SIGNAL (2) <-- Reason for the process abort
Last heartbeat 46.88 secs ago
System image name: switchh-dk9.5.2.1.SM15.0.1.bin
System image version: 5.2(1)SM1(5.1)

PID: 3207
Exit code: signal 6 (core dumped) <-- Indicates that a cores for the process was
generated.

CWD: /var/sysmgr/work
...
```

To check the redundancy status, enter this command:

- **show system redundancy status**

```
switch# show system redundancy status
Redundancy role
---------------
      administrative:   primary <-- Configured redundancy role
         operational:   primary <-- Current operational redundancy role

Redundancy mode
---------------
      administrative:   HA
         operational:   HA

This supervisor (sup-1)
-----------------------
```

```
    Redundancy state:   Active <-- Redundancy state of this VSM
    Supervisor state:   Active
      Internal state:   Active with HA standby

Other supervisor (sup-2)
------------------------
    Redundancy state:   Standby <-- Redundancy state of the other VSM
    Supervisor state:   HA standby
      Internal state:   HA standby <-- The standby VSM is in HA mode and in sync
```

To check the system internal redundancy status, enter this command:

- **show system internal redundancy info**

```
switch# show system internal redundancy info
My CP:
  slot: 0
  domain: 184 <-- Domain id used by this VSM
  role:   primary <-- Redundancy role of this VSM
  status: RDN_ST_AC <-- Indicates redundancy state (RDN_ST) of the this VSM is Active
(AC)
  state:  RDN_DRV_ST_AC_SB
  intr:   enabled
  power_off_reqs: 0
  reset_reqs:     0
Other CP:
  slot: 1
  status: RDN_ST_SB <-- Indicates redundancy state (RDN_ST) of the other VSM is
Standby (SB)
  active: true
  ver_rcvd: true
  degraded_mode: false <-- When true, it indicates that communication through the
control interface is faulty
Redun Device 0: <-- This device maps to the control interface
  name: ha0
  pdev: ad7b6c60
  alarm: false
  mac: 00:50:56:b7:4b:59
  tx_set_ver_req_pkts:   11590
  tx_set_ver_rsp_pkts:   4
  tx_heartbeat_req_pkts: 442571
  tx_heartbeat_rsp_pkts: 6
  rx_set_ver_req_pkts:   4
  rx_set_ver_rsp_pkts:   1
  rx_heartbeat_req_pkts: 6
  rx_heartbeat_rsp_pkts: 442546 <-- Counter should be increasing, as this indicates
that communication between VSM is working properly.
rx_drops_wrong_domain: 0
  rx_drops_wrong_slot:   0
  rx_drops_short_pkt:    0
  rx_drops_queue_full:   0
  rx_drops_inactive_cp:  0
  rx_drops_bad_src:      0
  rx_drops_not_ready:    0
  rx_drops_wrong_ver:    0
  rx_unknown_pkts:       0
Redun Device 1: <-- This device maps to the mgmt interface
  name: ha1
  pdev: ad7b6860
  alarm: true
  mac: ff:ff:ff:ff:ff:ff
  tx_set_ver_req_pkts:   11589
  tx_set_ver_rsp_pkts:   0
  tx_heartbeat_req_pkts: 12
```

```
  tx_heartbeat_rsp_pkts: 0
  rx_set_ver_req_pkts:   0
  rx_set_ver_rsp_pkts:   0
  rx_heartbeat_req_pkts: 0
  rx_heartbeat_rsp_pkts: 0 <-- When communication between VSM through the control
interface is interrupted but continues through the mgmt interface, the
rx_heartbeat_rsp_pkts will increase.
  rx_drops_wrong_domain: 0
  rx_drops_wrong_slot:   0
  rx_drops_short_pkt:    0
  rx_drops_queue_full:   0
  rx_drops_inactive_cp:  0
  rx_drops_bad_src:      0
  rx_drops_not_ready:    0
  rx_drops_wrong_ver:    0
  rx_unknown_pkts:       0
```

To check the system internal sysmgr state, enter this command:

- **show system internal sysmgr state**

```
switch# show system internal sysmgr state

The master System Manager has PID 1988 and UUID 0x1.
Last time System Manager was gracefully shutdown.
The state is SRV_STATE_MASTER_ACTIVE_HOTSTDBY entered at time Tue Apr 28 13:09:13
2009.

The '-b' option (disable heartbeat) is currently disabled.

The '-n' (don't use rlimit) option is currently disabled.

Hap-reset is currently enabled.

Process restart capability is currently disabled.

Watchdog checking is currently disabled.

Watchdog kgdb setting is currently enabled.


        Debugging info:

The trace mask is 0x00000000, the syslog priority enabled is 3.
The '-d' option is currently disabled.
The statistics generation is currently enabled.


        HA info:

slotid = 1    supid = 0
cardstate = SYSMGR_CARDSTATE_ACTIVE .
cardstate = SYSMGR_CARDSTATE_ACTIVE (hot switchover is configured enabled).
Configured to use the real platform manager.
Configured to use the real redundancy driver.
Redundancy register: this_sup = RDN_ST_AC, other_sup = RDN_ST_SB.
EOBC device name: eth0.
Remote addresses:  MTS - 0x00000201/3      IP - 127.1.1.2
MSYNC done.
Remote MSYNC not done.
Module online notification received.
Local super-state is: SYSMGR_SUPERSTATE_STABLE
Standby super-state is: SYSMGR_SUPERSTATE_STABLE
Swover Reason : SYSMGR_SUP_REMOVED_SWOVER <-- Reason for the last switchover
```

```
Total number of Switchovers: 0 <-- Number of switchovers
                             >> Duration of the switchover would be listed, if any.
Swover threshold settings: 20 switchovers within 1200 seconds
Switchovers within threshold interval: 0
Last switchover time: 0 seconds after system start time
Cumulative time between last 0 switchovers: 0
Start done received for 2 plugins, Total number of plugins = 2

        Statistics:

Message count:           0
Total latency:           0              Max latency:            0
Total exec:              0              Max exec:               0
```

To reload a module, enter this command:

- **reload module**

  ```
  switch# reload module 2
  ```

  This command reloads the secondary VSM.

Note Entering the **reload** command without specifying a module reloads the whole system.

To attach to the standby VSM console, enter this command:

- **attach module**

  The standby VSM console is not accessible externally but can be accessed from the active VSM through the **attach module** *module-number* command.

  ```
  switch# attach module 2
  ```

  This command attaches to the console of the secondary VSM.