Cisco Nexus 7000 Series NX-OS High Availability and Redundancy Guide

First Published: 2016-12-23

Last Modified: 2022-08-04

Americas Headquarters
Cisco Systems, Inc.
170 West Tasman Drive
San Jose, CA 95134-1706
USA
http://www.cisco.com
Tel: 408 526-4000
800 553-NETS (6387)
Fax: 408 527-0883

CONTENTS

Preface

The preface contains the following sections:

Preface

This preface describes the audience, organization, and conventions of the Book Title. It also provides information on how to obtain related documentation.

This chapter includes the following topics:

Audience

This publication is for experienced network administrators who configure and maintain Cisco NX-OS on Cisco Nexus 7000 Series Platform switches.

Document Conventions

Note

- As part of our constant endeavor to remodel our documents to meet our customers' requirements, we have modified the manner in which we document configuration tasks. As a result of this, you may find a deviation in the style used to describe these tasks, with the newly included sections of the document following the new format.
- The Guidelines and Limitations section contains general guidelines and limitations that are applicable to all the features, and the feature-specific guidelines and limitations that are applicable only to the corresponding feature.

Command descriptions use the following conventions:

Convention	Description
bold	Bold text indicates the commands and keywords that you enter literally as shown.
Italic	Italic text indicates arguments for which the user supplies the values.

Convention	Description
[x]	Square brackets enclose an optional element (keyword or argument).
[x | y]	Square brackets enclosing keywords or arguments separated by a vertical bar indicate an optional choice.
{x | y}	Braces enclosing keywords or arguments separated by a vertical bar indicate a required choice.
[x {y | z}]	Nested set of square brackets or braces indicate optional or required choices within optional or required elements. Braces and a vertical bar within square brackets indicate a required choice within an optional element.
`variable`	Indicates a variable for which you supply values, in context where italics cannot be used.
string	A nonquoted set of characters. Do not use quotation marks around the string or the string will include the quotation marks.

Examples use the following conventions:

Convention	Description
`screen font`	Terminal sessions and information the switch displays are in screen font.
`boldface screen font`	Information you must enter is in boldface screen font.
italic screen font	Arguments for which you supply values are in italic screen font.
< >	Nonprinting characters, such as passwords, are in angle brackets.
[ ]	Default responses to system prompts are in square brackets.
!, #	An exclamation point (!) or a pound sign (#) at the beginning of a line of code indicates a comment line.

This document uses the following conventions:

Note Means *reader take note*. Notes contain helpful suggestions or references to material not covered in the manual.

Caution Means *reader be careful*. In this situation, you might do something that could result in equipment damage or loss of data.

Related Documentation

Documentation for Cisco Nexus 7000 Series Switches is available at:

- Configuration Guides

http://www.cisco.com/c/en/us/support/switches/nexus-7000-series-switches/products-installation-and-configuration-guides-list.html

- Command Reference Guides

  http://www.cisco.com/c/en/us/support/switches/nexus-7000-series-switches/products-command-reference-list.html

- Release Notes

  http://www.cisco.com/c/en/us/support/switches/nexus-7000-series-switches/products-release-notes-list.html

- Install and Upgrade Guides

  http://www.cisco.com/c/en/us/support/switches/nexus-7000-series-switches/products-installation-guides-list.html

- Licensing Guide

  http://www.cisco.com/c/en/us/support/switches/nexus-7000-series-switches/products-licensing-information-listing.html

Documentation for Cisco Nexus 7000 Series Switches and Cisco Nexus 2000 Series Fabric Extenders is available at the following URL:

http://www.cisco.com/c/en/us/support/switches/nexus-2000-series-fabric-extenders/products-installation-and-configuration-guides-list.html

Documentation Feedback

To provide technical feedback on this document, or to report an error or omission, please send your comments to nexus7k-docfeedback@cisco.com. We appreciate your feedback.

Communications, Services, and Additional Information

- To receive timely, relevant information from Cisco, sign up at Cisco Profile Manager.
- To get the business impact you're looking for with the technologies that matter, visit Cisco Services.
- To submit a service request, visit Cisco Support.
- To discover and browse secure, validated enterprise-class apps, products, solutions and services, visit Cisco Marketplace.
- To obtain general networking, training, and certification titles, visit Cisco Press.
- To find warranty information for a specific product or product family, access Cisco Warranty Finder.

Cisco Bug Search Tool

Cisco Bug Search Tool (BST) is a web-based tool that acts as a gateway to the Cisco bug tracking system that maintains a comprehensive list of defects and vulnerabilities in Cisco products and software. BST provides you with detailed defect information about your products and software.

CHAPTER 1

New and Changed Information

New and Changed Information

The table below summarizes the new and changed features for this document and shows the releases in which each feature is supported. Your software release might not support all the features in this document. For the latest caveats and feature information, see the Bug Search Tool and the release notes for your software release.

Table 1: New and Changed Information

Feature	Description	Changed in Release
Supervisor-to-Supervisor EOBC Link Redundancy	This feature provides redundancy for EOBC communication between the active and standby supervisors by enabling the secondary redundant EOBC link in case the primary EOBC link fails and vice-versa.	8.4(1)
Internal CRC Error Detection and Isolation	This feature enables the switches to detect CRC errors that occur internally within a switch and isolate the source of these errors when the configured threshold value is reached.	8.4(1)
Message and Transaction Service	The System Message Logging contains new logs that indicates the highest MTS memory users. Additionally, detailed memory usage stats with timestamps are collected per application.	8.0(1)

CHAPTER 2

Overview

Cisco NX-OS is a resilient operating system that is specifically designed for high availability at the network, system, and process level.

This chapter describes high availability (HA) concepts and features for Cisco NX-OS devices and includes the following sections:

Licensing Requirements

For a complete explanation of Cisco NX-OS licensing recommendations and how to obtain and apply licenses, see the *Cisco NX-OS Licensing Guide*.

Information About High Availability

To prevent or minimize traffic disruption during hardware or software failures, Cisco NX-OS has these features:

- Redundancy—Cisco NX-OS HA provides physical and software redundancy at every component level, spanning across the physical, environmental, power, and system software aspects of its architecture.
- Isolation of planes and processes—Cisco NX-OS HA provides isolation between control and data forwarding planes within the device and between software components, so that a failure within one plane or process does not disrupt others.
- 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—The Nexus 7000 series supports an active and standby dual supervisor configuration. State and configuration remain constantly synchronized between the two supervisor modules to provide seamless and stateful switchover in the event of a supervisor module failure.
- Nondisruptive upgrades—Cisco NX-OS supports the in-service software upgrade (ISSU) feature, which allows you to upgrade the device software while the switch continues to forward traffic. ISSU reduces or eliminates the downtime typically caused by software upgrades.

Service-Level High Availability

Cisco NX-OS has a modularized architecture that compartmentalizes components for fault isolation, redundancy, and resource efficiency.

For additional details about service-level HA, see chapter *Service-Level High Availability*.

Isolation of Processes

In the Cisco NX-OS software, independent processes, known as*services,* perform a function or set of functions for a subsystem or feature set. Each service and service instance runs as an independent, protected process. This approach provides a highly fault-tolerant software infrastructure and fault isolation between services. A failure in a service instance (such as 802.1Q) will not affect any other services running at that time (such as the Link Aggregation Control Protocol [LACP]). In addition, each instance of a service can run as an independent process, which means that two instances of a routing protocol (for example, two instances of the Open Shortest Path First [OSPF] protocol) can run as separate processes.

Process Restartability

Cisco NX-OS processes run in a protected memory space independently from each other and the kernel. This process isolation provides fault containment and enables rapid restarts. Process restartability ensures that process-level failures do not cause system-level failures. In addition, most services can perform stateful restarts, which allows a service that experiences a failure to be restarted and to resume operations transparently to other services within the platform and to neighboring devices within the network.

System-Level High Availability

The Nexus 7000 series is protected from system failure by redundant hardware components and a high-availability software framework.

For additional information about system-level HA features, see chapter, *System-Level High Availability*.

Physical Redundancy

The Nexus 7000 series has the following physical redundancies:

- Power Supply Redundancy—The Cisco Nexus 7000 series chassis supports three power supply modules on a Cisco Nexus 7010 switch and up to four power supplies on a Cisco Nexus 7018 switch, each of which is composed of two internalized isolated power units, giving it two power paths per modular power supply, and six paths in total, per chassis, when fully populated.

- Fan Tray Redundancy—The Cisco Nexus 7010 chassis contains two redundant system fan trays for I/O module cooling and two redundant fan trays for switch fabric module cooling. One of each pair of fan trays is sufficient to provide system cooling. There is no time limit for replacing a failed Cisco Nexus 7010 fan tray, but to ensure the proper airflow, you must leave the failed fan tray in place.

  The Cisco Nexus 7018 chassis contains two fan trays, each of which is required to cool the modules in the chassis. The upper fan tray cools slots 1 to 9 and the fabric modules. The lower fan tray cools slots 10 to 18. Each of these fan trays is hot swappable, but you must replace a fan tray within 3 minutes of removal or the switch will shut down.

- Fabric Redundancy—Cisco NX-OS provides switching fabric availability through redundant switch fabric modules. You can configure a single Cisco Nexus 7000 series chassis with one to five switch fabric cards for capacity and redundancy. Each I/O module installed in the system automatically connects to and uses all functionally installed switch fabric modules. A failure of a switch fabric module triggers an automatic reallocation and balancing of traffic across the remaining active switch fabric modules. Replacing the failed fabric module reverses this process. When you insert the fabric module and bring it online, traffic is again redistributed across all installed fabric modules and redundancy is restored.

- Supervisor Module Redundancy—The Cisco Nexus 7000 series chassis supports dual supervisor modules to provide redundancy for the control and management plane. A dual supervisor configuration operates in an active/standby capacity in which only one of the supervisor modules is active at any given time, while the other acts as a standby backup. The state and configuration remain constantly synchronized between the two supervisor modules to provide a statefu1 switchover if the active supervisor module fails.

ISSU

Cisco NX-OS allows you to perform an in-service software upgrade (ISSU), which is also known as a nondisruptive upgrade. The modular software architecture of Cisco NX-OS supports plug-in-based services and features, which allow you to perform complete image upgrades of supervisors and switching modules with little to no impact on other modules. Because of this design, you can upgrade Cisco NX-OS nondisruptively with no impact to the data forwarding plane and allow for nonstop forwarding during a software upgrade, even between full image versions.

For additional details about ISSU, see ISSU and High Availability, on page 43.

VDCs

Cisco NX-OS implements a logical virtualization at the device level, which allows multiple instances of a device to operate on the same physical switch simultaneously. These logical operating environments are known as*virtual device contexts*, or VDCs. VDCs provide logically separate device environments that you can independently configure and manage. This degree of isolation provides fault isolation in addition to security and administrative benefits. Human error or failure conditions occur when the configuration is isolated within a given virtual device. While virtual device contexts are not primarily a high-availability feature, the operationally independent fault domains contribute to availability and prevent service disruptions that are associated with device configuration.

For more information on VDCs, see *Cisco Nexus 7000 Series NX-OS Virtual Device Context Configuration Guide*.

Network-Level High Availability

Network convergence is optimized by providing tools and functions to make both failover and fallback transparent and fast.

For additional information about network-level HA features, see *Network-Level High Availability.*.

Internal CRC Error Detection and Isolation

Starting with the Cisco NX-OS Release 8.4(1), the Internal Cyclic Redundancy Check (CRC) error detection and isolation feature is supported on the Cisco Nexus 7000 Series switches.

This feature enables the switches to detect CRC errors that occur internally within a switch and isolate the source of these errors when the configured threshold value is reached. Use the **hardware fabric crc threshold** command to enable this feature. This command also enables powering down of any switching or supervisor module that exceeds the internal CRC error threshold limit. By default, the Internal CRC Error Detection and Isolation feature is disabled.

Detection of CRC errors along with powering down of any switching or supervisor module that exceeds the internal CRC error threshold limit is only supported on devices in which SM15 or SAC ASICs are present. No action is triggered on devices on which any other ASICs are present.

Internal CRC errors are usually caused by a fault in the system. Such faults may be transient, such as an ungracefully removed module, or permanent, such as a badly seated module, or, in rare cases, a failing or failed hardware component. Module refers either a switching module or a supervisor module. The rate of errors depends on many factors and may range from very high to very low. The error-rate threshold is configurable as a system-wide value, but separate error counts are maintained for each module to identify an error source. Errors on each module are handled individually when the error count exceeds the threshold.

Polling of internal CRC errors is based on an interrupt timer. In case an interrupt flag is set, it indicates that a CRC error has occured and the CRC counter is correspondingly increased. When the CRC counter value crosses the threshold value, the CRC error handler is triggered. The duration of the interrupt timer depends on factors such as the supervisor module being used and the number of ASICs in the device.

Note The counters are reset at 24 hours from the time the Internal Cyclic Redundancy Check (CRC) detection and isolation feature was first configured.

Configuring Internal CRC Error Detection and Isolation

Step 1 Enter global configuration mode:

switch# **configure terminal**

Step 2 Enable internal CRC error detection and isolation:

switch(config)# **hardware fabric crc threshold** *<threshold-count>*

Note The error rate is measured over sequential 24-hour window. The range of threshold is 1 to 100. If the threshold is not specified, the default threshold value is 3.

Step 3 (Optional) Disable internal CRC error detection and isolation:

switch(config)# **no hardware fabric crc**

Step 4 (Optional) Save the running configuration to the startup configuration:

switch(config)# **copy running-config startup-config**

Running Configuration

This example shows how to enable internal CRC error detection and isolation.

```
configure terminal
 hardware fabric crc threshold 5
```

Additional Management Tools for Availability

Cisco NX-OS incorporates several Cisco system management tools for monitoring and notification of system availability events.

GOLD

Cisco Generic On-Line Diagnostics (GOLD) subsystem and additional monitoring processes on the supervisor facilitate the triggering of a stateful failover to the redundant supervisor upon the detection of unrecoverable critical failures, service restartability errors, kernel errors, or hardware failures.

For information about configuring GOLD, see the *Cisco Nexus 7000 Series NX-OS System Management Configuration Guide*.

EEM

Cisco Embedded Event Manager (EEM) consists of Event Detectors, the Event Manager, and an Event Manager Policy Engine. Using EEM, you can define policies to take specific actions when the system software recognizes certain events through the Event Detectors. The result is a flexible set of tools to automate many network management tasks and to direct the operation of Cisco NX-OS to increase availability, collect information, and notify external systems or personnel about critical events.

For information about configuring EEM, see the *Cisco Nexus 7000 Series NX-OS System Management Configuration Guide*.

Smart Call Home

Combining Cisco GOLD and Cisco EEM capabilities, Smart Call Home provides an e-mail-based notification of critical system events. Smart Call Home has message formats that are compatible with pager services, standard e-mail, or XML-based automated parsing applications. You can use this feature to page a network support engineer, e-mail a network operations center, or use Cisco Smart Call Home services to automatically generate a case with Cisco's Technical Assistance Center (TAC).

For information about configuring Smart Call Home, see the *Cisco Nexus 7000 Series NX-OS System Management Configuration Guide*.

CHAPTER 3

Service-Level High Availability

This chapter describes the Cisco NX-OS service restartability for service-level HA and includes the following sections:

Information About Cisco NX-OS Service Restarts

The Cisco NX-OS service restart features allows you to restart a faulty service without restarting the supervisor to prevent process-level failures from causing system-level failures. You can restart a service depending on current errors, failure circumstances, and the high-availability policy for the service. A service can undergo either a stateful or stateless restart. Cisco NX-OS allows services to store run-time state information and messages for a stateful restart. In a stateful restart, the service can retrieve this stored state information and resume operations from the last checkpoint service state. In a stateless restart, the service can initialize and run as if it had just been started with no prior state.

Not all services are designed for stateful restart. For example, Cisco NX-OS does not store run-time state information for Layer 3 routing protocols (such as Open Shortest Path First [OSPF] and Routing Information Protocol [RIP]). Their configuration settings are preserved across a restart, but these protocols are designed to rebuild their operational state using information obtained from neighbor routers. For details on the high-availability functionality of Layer 3 protocols, see chapter, *Network-Level High Availability* .

Virtualization Support

For complete information on VDCs, see the *Cisco Nexus 7000 Series NX-OS Virtual Device Context Configuration Guide*.

Restartability Infrastructure

Cisco NX-OS allows stateful restarts of most processes and services. Back-end management and orchestration of processes, services, and applications within a platform are handled by a set of high-level system-control services.

System Manager

The System Manager directs overall system function, service management, and system health monitoring, and enforces high-availability policies. The System Manager is responsible for launching, stopping, monitoring, restarting services, and initiating and managing the synchronization of service states and supervisor states for stateful switchover.

Persistent Storage Service

Cisco NX-OS services use the persistent storage service (PSS) to store and manage the operational run-time information. The PSS component works with system services to recover states in the event of a service restart. PSS functions as a database of state and run-time information, which allows services to make a checkpoint of their state information whenever needed. A restarting service can recover the last known operating state that preceded a failure, which allows for a stateful restart.

Each service that uses PSS can define its stored information as private (it can be read only by that service) or shared (the information can be read by other services). If the information is shared, the service can specify that it is local (the information can be read only by services on the same supervisor) or global (it can be read by services on either supervisor or on modules). For example, if the PSS information of a service is defined as shared and global, services on other modules can synchronize with the PSS information of the service that runs on the active supervisor.

Message and Transaction Service

The message and transaction service (MTS) is a high-performance interprocess communications (IPC) message broker that specializes in high-availability semantics. MTS handles message routing and queuing between services on and across modules and between supervisors. MTS facilitates the exchange of messages such as event notification, synchronization, and message persistency between system services and system components. MTS can maintain persistent messages and logged messages in queues for access even after a service restart.

MTS provides extensive serviceability features. For example, MTS provides notifications to inform an application when its queue has reached a predefined limitation. Corresponding to each notification, a default callback action is defined in MTS. From Cisco NX-OS Release 8.0(1), the System Message Logging contains new logs that indicates the highest MTS memory users. These logs are set to the severity level 4. Additionally, detailed memory usage stats with timestamps are collected per application. If an application contains an issue, you can use the command **show sys int mts sup sap APP_SAP_NUM queue_stats** to collect the tech support.

```
MTS-4-MTS_MEM_USAGE: AT 886890 usecs after 11/16/2016 15:19:27: Total mem used: 76048384
(45%)
```

```
MTS-4-MTS_MEM_USAGE: top user 1: Node=0x101 vdc=1 sap=1(SRVDESC_MTS_SYNC_THREAD)
bytes=77724160 msgs=9421
MTS-4-MTS_MEM_USAGE: top user 2: Node=0x109 vdc=3 sap=63(mping_server) bytes=1736704 msgs=53
```

Explanation: MTS total memory usage and top memory users. This information will assist in debugging application issues.

Recommended Action: None.

HA Policies

Cisco NX-OS allows each service to have an associated set of internal HA policies that define how a failed service will be restarted. Each service can have four defined policies—a primary and secondary policy when two supervisors are present, and a primary and secondary policy when only one supervisor is present. If no HA policy is defined for a service, the default HA policy to be performed upon a service failure will be a switchover if two supervisors are present, or a supervisor reset if only one supervisor is present.

Each HA policy specifies three parameters:

- Action to be performed by the System Manager:
  - Stateful restart
  - Stateless restart
  - Supervisor switchover (or restart)
- Maximum retries—Specifies the number of restart attempts to be performed by the System Manager. If the service has not restarted successfully after this number of attempts, the HA policy is considered to have failed, and the next HA policy is used. If no other HA policy exists, the default policy is applied, resulting in a supervisor switchover or restart.
- Minimum lifetime—Specifies the time that a service must run after a restart attempt to consider the restart attempt as successful. The minimum lifetime is no less than four minutes.

Process Restartability

Process restartability ensures that a failed service can recover and resume operations without disrupting the data plane or other services. Depending on the service HA policies, previous restart failures, and the health of other services running on the same supervisor, the System Manager determines the action to be taken when a service fails.

The action taken by the System Manager for various failure conditions is described in the following table:

Table 2: System Manager Action for Various Failure Cases

Failure	
Service/process exception	Service restart
Service/process crash	Service restart
Unresponsive service/process	Service restart

Failure	
Repeated service failure	Supervisor reset (single) or switchover (dual)
Unresponsive System Manager	Supervisor reset (single) or switchover (dual)
Supervisor hardware failure	Supervisor reset (single) or switchover (dual)
Kernel failure	Supervisor reset (single) or switchover (dual)
Watchdog timeout	Supervisor reset (single) or switchover (dual)

Types of Process Restarts

A failed service is restarted by one of the methods described in this section, depending on the service's HA implementation and HA policies,

Stateful Restarts

When a restartable service fails, it is restarted on the same supervisor. If the new instance of the service determines that the previous instance was abnormally terminated by the operating system, the service then determines whether a persistent context exists. The initialization of the new instance attempts to read the persistent context to build a run-time context that makes the new instance appear like the previous one. After the initialization is complete, the service resumes the tasks that it was performing when it stopped. During the restart and initialization of the new instance, other services are unaware of the service failure. Any messages that are sent by other services to the failed service are available from the MTS when the service resumes.

Whether or not the new instance survives the stateful initialization depends on the cause of failure of the previous instance. If the service is unable to survive a few subsequent restart attempts, the restart is considered as failed. In this case, the System Manager performs the action specified by the HA policy of the services, forcing either a stateless restart, no restart, or a supervisor switchover or reset.

During a successful stateful restart, there is no delay while the system reaches a consistent state. Stateful restarts reduce the system recovery time after a failure.

The events before, during, and after a stateful restart are as follows:

1. The running services make a checkpoint of their run-time state information to the PSS.
2. The System Manager monitors the health of the running services that use heatbeats.
3. The System Manager restarts a service instantly when it crashes or hangs.
4. After restarting, the service recovers its state information from the PSS and resumes all pending transactions.
5. If the service does not resume a stable operation after multiple restarts, the System Manager initiates a reset or switchover of the supervisor.
6. Cisco NX-OS collects the process stack and core for debugging purposes with an option to transfer core files to a remote location.

When a stateful restart occurs, Cisco NX-OS sends a syslog message of level LOG_ERR. If SNMP traps are enabled, the SNMP agent sends a trap. If the Smart Call Home service is enabled, the service sends an event message.

Stateless Restarts

Cisco NX-OS infrastructure components manage stateless restarts. During a stateless restart, the System Manager identifies the failed process and replaces it with a new process. The service that failed does not maintain its run-time state upon the restart. The service can either build the run-time state from the running configuration, or if necessary, exchange information with other services to build a run-time state.

When a stateless restart occurs, Cisco NX-OS sends a syslog message of level LOG_ERR. If SNMP traps are enabled, the SNMP agent sends a trap. If the Smart Call Home service is enabled, the service sends an event message.

Switchovers

If a standby supervisor is available, Cisco NX-OS performs a supervisor switchover rather than a supervisor restart whenever multiple failures occur at the same time, because these cases are considered unrecoverable on the same supervisor. For example, if more than one HA application fails, that is considered an unrecoverable failure. For detailed information about supervisor switchovers and restarts, see chapter, *System level high availability*.

Restarts on Standby Supervisor Services

When a service fails on a supervisor that is in the standby state, the System Manager does not apply the HA policies and restarts the service after a delay of 30 seconds. The delay ensures that the active supervisor is not overloaded by repeated standby service failures and synchronizations. If the service being restarted requires synchronization with a service on the active supervisor, the standby supervisor is taken out of hot standby mode until the service is restarted and synchronized. Services that are not restartable cause the standby supervisor to reset.

When a standby service restart occurs, Cisco NX-OS sends a syslog message of level LOG_ERR. If SNMP traps are enabled, the SNMP agent sends a trap. If the Smart Call Home service is enabled, the service sends an event message.

Restarts on Switching Module Services

When services fail on a switching module or another nonsupervisor module, the recovery action is determined by HA policies for those services. Because service failures on nonsupervisor module services do not require a supervisor switchover, the recovery options are a stateful restart, a stateless restart, or a module reset. If the module is capable of a nondisruptive upgrade, it is also capable of a nondisruptive restart.

When a nonsupervisor module service restart occurs, Cisco NX-OS sends a syslog message of level LOG_ERR. If SNMP traps are enabled, the SNMP agent sends a trap. If the Smart Call Home service is enabled, the service sends an event message.

Restarts on Services Within a VDC

When a service fails and all HA policies have been unsuccessful in restarting the service, the next action is typically a supervisor restart or switchover, However, if the service is running within a VDC, a VDC policy can specify that a restart of the VDC will be attempted before a supervisor restart or switchover.

For more information on VDCs, see the *Cisco Nexus 7000 Series NX-OS Virtual Device Context Configuration Guide*.

Troubleshooting Restarts

When a service fails, the system generates information that can be used to determine the cause of the failure. The following sources of information are available:

- Every service restart generates a syslog message of level LOG_ERR.
- If the Smart Call Home service is enabled, every service restart generates a Smart Call Home event.
- If SNMP traps are enabled, the SNMP agent sends a trap when a service is restarted,
- When a service failure occurs on a local module, you can view a log of the event by entering the **show processes log** command in that module. The process logs are persistent across supervisor switchovers and resets.
- When a service fails, a system core image file is generated. You can view recent core images by using the **show cores** command on the active supervisor. Core files are not persistent across supervisor switchovers and resets, but you can configure the system to export core files to an external server by using a file transfer utility such as the Trivial File Transfer Protocol (TFTP).
- CISCO-SYSTEM-MIB contains a table for cores (cseSwCoresTable).

For information on collecting and using the generated information relating to service failures, see the *Cisco Nexus 7000 Series NX-OS Troubleshooting Guide*.

Related Documents

Related Topic	Document Title
Virtual device context (VDC)	*Cisco Nexus 7000 Series NX-OS Virtual Device Context Configuration Guide*
Supervisor switchovers	*System-Level High Availability*
Troubleshooting	*Cisco Nexus 7000 Series NX-OS Troubleshooting Guide*
Cisco NX-OS fundamentals	*Cisco Nexus 7000 Series NX-OS Fundamentals Configuration Guide*
Licensing	*Cisco Nexus 7000 Series NX-OS Licensing Guide*

Standards

Standards	Title
No new or modified standards are supported by this feature, and support for existing standards has not been modified by this feature.	—

MIBs

MIBs	MIBs Link
• CISCO-SYSTEM-EXT-MIB: ciscoHaGroup, cseSwCoresTable, cseHaRestartNotify, cseShutDownNotify, cseFailSwCoreNotify, cseFailSwCoreNotifyExtended • CISCO-PROCESS-MIB • CISCO-RF-MIB	To locate and download MIBs, go to the following URL: https://cfnng.cisco.com/mibs

RFCs

RFCs	Title
No RFCs are supported by this feature	—

Technical Assistance

Description	Link
Technical Assistance Center (TAC) home page, containing 30,000 pages of searchable technical content, including links to products, technologies, solutions, technical tips, and tools. Registered Cisco.com users can log in from this page to access even more content.	http://www.cisco.com/cisco/web/support/index.html

Note This chapter refers to processes and services interchangably. A process is considered to be a running instance of a service.

CHAPTER 4

Network-Level High Availability

This chapter describes Cisco NX-OS network high availability and includes the following sections:

Information About Network-Level High Availability

Cisco NX-OS network-level HA is optimized by tools and functionality that provide failovers and fallbacks transparently and quickly. The features described in this chapter ensure high availability at the network level.

Virtualization Support

Each virtual device context (VDC) in a system runs a separate Spanning Tree Protocol (STP), which includes extensions to support virtualization. Each VDC can also run one or more instances of a routing protocol. The network-level HA features described in this chapter apply to a failure or restart of a VDC in the same manner as a failure or restart of the system.

Note For complete information on VDCs, see the *Cisco Nexus 7000 Series NX-OS Virtual Device Context Configuration Guide*.

Spanning Tree Protocol

Note Spanning Tree Protocol (STP)refers to IEEE 802.1w and IEEE 802.1s. If this publication is referring to the IEEE 802.1D STP, 802.1D is stated specifically.

When you create fault-tolerant internetworks, you must have a loop-free path between all nodes in a network. Multiple active paths between end stations cause loops in the network that result in network devices learning end station MAC addresses on multiple Layer 2 LAN ports. This condition can result in a broadcast storm, which creates an unstable network.

STP provides a loop-free network at the Layer 2 level. Layer 2 LAN ports send and receive STP frames, which are called Bridge Protocol Data Units (BPDUs), at regular intervals. Network devices do not forward these frames but use the frames to determine the network topology and to construct a loop-free path within that topology. Using the spanning tree topology, STP forces redundant data paths into a blocked state. If a network segment in the spanning tree fails and a redundant path exists, the STP algorithm recalculates the spanning tree topology and activates the blocked path.

Cisco NX-OS also supports Multiple Spanning Tree Protocol (MSTP.The multiple independent spanning tree topology enabled by MSTP provides multiple forwarding paths for data traffic, enables load balancing, and reduces the number of STP instances required to support a large number of VLANs.

MST incorporates Rapid Spanning Tree Protocol (RSTP), which allows rapid convergence. MST improves the fault tolerance of the network because a failure in one instance (forwarding path) does not affect other instances (forwarding paths).

Note You can configure spanning tree parameters only on Layer 2 interfaces; a spanning tree configuration is not allowed on a Layer 3 interface. For information on creating Layer 2 interfaces, see the *Cisco Nexus 7000 Series NX-OS Interfaces Configuration Guide*.

For details about STP behavior and configuration, see the *Cisco Nexus 7000 Series NX-OS Layer 2 Switching Configuration Guide*.

Virtual Port Channels

The major limitation in classic port channel communication is that the port channel operates only between two devices. In large networks, the support of multiple devices together is often a design requirement to provide some form of hardware failure alternate path. This alternate path is often connected in a way that would cause a loop, limiting the benefits gained with port channel technology to a single path. To address this limitation, Cisco NX-OS provides a technology called virtual port channel (vPC). Although a pair of switches acting as a vPC peer endpoint looks like a single logical entity to port channel-attached devices, the two devices that act as the logical port channel endpoint are still two separate devices. This environment combines the benefits of hardware redundancy with the benefits of port channel loop management.

For more information on vPCs, see the *Cisco Nexus 7000 Series NX-OS Interfaces Configuration Guide*.

First-Hop Redundancy Protocols

Within a group of two or more routers , first-hop redundancy protocols (FHRPs) allow a transparent failover of the first-hop IP router. Cisco NX-OS supports the following FHRPs:

- Hot Standby Router Protocol (HSRP)—HSRP provides first-hop routing redundancy for IP hosts on Ethernet networks configured with a default gateway IP address. An HSRP router group of two or more routers chooses an active gateway and a standby gateway. The active gateway routes packets while the standby gateway remains idle until the active gateway fails or when preset conditions are met.

  Many host implementations do not support any dynamic router discovery mechanisms but can be configured with a default router. Running a dynamic router discovery mechanism on every host is not feasible for a number of reasons, including administrative overhead, processing overhead, and security issues. HSRP provides failover services to these hosts.

- Virtual Router Redundancy Protocol (VRRP)—VRRP dynamically assigns responsibility for one or more virtual routers to the VRRP routers on a LAN, which allows several routers on a multi-access link to use the same virtual IP address. A VRRP router is configured to run VRRP with one or more other routers attached to a LAN. One router is elected as the virtual router master, while the other routers act as backups if the virtual router master fails.

- Gateway Load Balancing Protocol (GLBP)—GLBP provides path redundancy for IP by sharing protocol and Media Access Control (MAC) addresses between redundant gateways. In addition, GLBP allows a group of Layer 3 routers to share the load of the default gateway on a LAN. A GLBP router can automatically assume the forwarding function of another router in the group if the other router fails.

  GLBP performs a similar function to HSRP and the Virtual Router Redundancy Protocol (VRRP), which allows multiple routers to participate in a virtual group configured with a virtual IP address. GLBP performs an additional load balancing function that HSRP and VRRP do not provide. GLBP shares the forwarding load among all routers in a GLBP group instead of allowing a single router to handle the entire load while the other routers remain idle. HSRP and VRRP elect one member as the active router to forward packets to the virtual IP address for the group. The other routers in the group are redundant until the active router fails.

For configuration details about FHRPs, see the *Cisco Nexus 7000 Series NX-OS Unicast Routing Configuration Guide*.

Nonstop Forwarding in Routing Protocols

Cisco NX-OS provides a multilevel high-availability architecture. OSPFv2 supports stateful restart, which is also referred to as non-stop routing (NSR). If OSPFv2 experiences problems, it attempts to restart from its previous run time state. The neighbors would not register any neighbor event in this case.

If the first restart was not successful and another problem occurs, OSPFv2 attempts a graceful restart. A graceful restart, or nonstop forwarding (NSF), allows OSPFv2 to remain in the data forwarding path through a process restart. When OSPFv2 needs to do a graceful restart, it first sends a link-local opaque (type 9) LSA, called a grace LSA. (For more information about opaque LSAs, see the *Cisco Nexus 7000 Series NX-OS Unicast Routing Configuration Guide*.) The restarting of the OSPFv2 platform is called NSF capable. The grace LSA includes a grace period, which is a specified time that the neighbor OSPFv2 interfaces hold onto the LSAs from the restarting OSPFv2 interface. (Typically, OSPFv2 tears down the adjacency and discards all LSAs from a down or restarting OSPFv2 interface.) The participating neighbors, which are called NSF helpers, keep all LSAs that originate from the restarting OSPFv2 interface as if the interface were still adjacent. When the restarting OSPFv2 interface is operational again, it rediscovers its neighbors, establishes adjacency, and starts sending its LSA updates again. At this point, the NSF helpers recognize that graceful restart has finished.

Scenarios where a stateful restart is used:

- First recovery attempt after a process experiences problems.
- ISSU
- User-initiated switchover using the **system switchover** command.
- Active supervisor reload using the **reload module** <*active sup*> command.

Scenarios where graceful restart is used:

- Second recovery attempt after a process experiences problems within a 4 minute interval.
- Manual restart of the process using the **restart ospfv3** command.
- Active supervisor removal.

Note The Cisco Nexus 7000 series devices support the Internet Engineering Task Force (IETF) version only. As a result, NSF IETF must be explicitly configured under the routing protocols in the Virtual Switching System (VSS). Use the **nsf ietf** command in router configuration mode for NSF IETF configuration. No additional configuration is required on the Cisco Nexus 7000 pairs because they run NSF IETF graceful-restart by default. However, each neighbor device that will become Layer 3 adjacent must have NSF configured and the same mode of NSF must be enabled to successfully operate a graceful failover.

Related Documents

Related Topic	Document Title
Virtual device context (VDC)	*Cisco Nexus 7000 Series NX-OS Virtual Device Context Configuration Guide*
Graceful restart	*Cisco Nexus 7000 Series NX-OS Unicast Routing Configuration Guide*
In-service software upgrades (ISSU)	ISSU and High Availability, on page 43
Licensing	*Cisco Nexus 7000 Series NX-OS Licensing Guide*

Standards

Standards	Title
IEEE 802.1Q-2006 (formerly known as IEEE 802.1s), IEEE 802.1D-2004 (formerly known as IEEE 802.1w), IEEE 802.1D, IEEE 802.1t	—

MIBs

MIBs	MIBs Link
• CISCO-SYSTEM-EXT-MIB: ciscoHaGroup, cseSwCoresTable, cseHaRestartNotify, cseShutDownNotify, cseFailSwCoreNotify, cseFailSwCoreNotifyExtended • CISCO-STP-EXTENSION-MIB • CISCO-PROCESS-MIB • CISCO-RF-MIB	To locate and download MIBs, go to the following URL: https://cfnng.cisco.com/mibs

RFCs

RFCs	Title
No RFCs are supported by this feature	—

Technical Assistance

Description	Link
Technical Assistance Center (TAC) home page, containing 30,000 pages of searchable technical content, including links to products, technologies, solutions, technical tips, and tools. Registered Cisco.com users can log in from this page to access even more content.	http://www.cisco.com/cisco/web/support/index.html

CHAPTER 5

System-Level High Availability

This chapter describes the Cisco NX-OS HA system and application restart operations and includes the following sections:

Information About Cisco NX-OS System-Level High Availability

Cisco NX-OS system-level HA mitigates the impact of hardware or software failures and is supported by the following features:

- Redundant hardware components:
  - Supervisor
  - Switch fabric
  - Power supply
  - Fan trays

  For details about physical requirements and redundant hardware components, respectively, see the *Cisco Nexus 7000 Series Site Preparation Guide* and the *Cisco Nexus 7000 Series Hardware Installation and Reference Guide*.

- HA software features:
  - For details about configuring and performing nondisruptive upgrades, see ISSU and High Availability, on page 43.

- Nonstop forwarding (NSF) — For details about nonstop forwarding, also known as graceful restart, see the *Cisco Nexus 7000 Series NX-OS Unicast Routing Configuration Guide*.
- Virtual device contexts (VDCs) — For details about VDCs and HA, see the *Cisco Nexus 7000 Series NX-OS Virtual Device Context Configuration Guide*.
- Generic online diagnostics (GOLD) — For details about configuring GOLD, see the *Cisco Nexus 7000 Series NX-OS System Management Configuration Guide*.
- Embedded event manager (EEM) — For details about configuring EEM, see the *Cisco Nexus 7000 Series NX-OS System Management Configuration Guide*.
- Smart Call Home — For details about configuring Smart Call Home, see the *Cisco Nexus 7000 Series NX-OS System Management Configuration Guide*.

Virtualization Support

For information about system-level high availability within a virtual device context (VDC), see, chapter *Network-level High Availability*.

Note For complete information on VDCs, see the *Cisco Nexus 7000 Series NX-OS Virtual Device Context Configuration Guide*.

Physical Redundancy

The Nexus 7000 series includes the following physical redundancies:

For additional details about physical redundancies, see the *Cisco Nexus 7000 Series Site Preparation Guide* and the *Cisco Nexus 7000 Series Hardware Installation and Reference Guide*.

Power Supply Redundancy

The Nexus 7000 series supports up to three power supply modules on a Cisco Nexus 7010 switch and up to four power supplies on a Cisco Nexus 7018 switch. Each power supply module can deliver up to 7.5 KW, depending on the number of inputs and the input line voltage. By installing two or three modules, you can ensure that the failure of one module will not disrupt system operations. You can replace the failed module while the system is operating. For information on power supply module installation and replacement, see the *Cisco Nexus 7000 Series Hardware Installation and Reference Guide*.

For further redundancy, each power supply module includes two internalized isolated power units, which give it two power paths per modular power supply, and six paths in total, per chassis, when fully populated. In addition, the power subsystem allows the three power supplies to be configured in any one of four redundancy modes.

Power Modes

Each of the four available power redundancy modes imposes different power budgeting and allocation models, which in turn deliver varying usable power yields and capacities. For more information regarding power budgeting, usable capacity, planning requirements, and redundancy configuration, see the *Cisco Nexus 7000 Series Hardware Installation and Reference Guide*.

The redundancy modes are only for software allocation of power to the system. In all the modes, the power supplies will be load shared based on their input and functionality. The available power in the system is determined at the start by the number of supplies in the system.

The available power supply redundancy modes are described in the Table below.

Table 3: Power Redundancy Modes

Redundancy Mode	Description
Combined	This mode does not provide power redundancy. The available power is the total power capacity of all power supplies.
insrc-redundant	This mode utilizes two electrical grids, each one powering a half module within each power supply. If one power grid goes down, each power supply continues to draw power through its other half module. The available power is the amount of power by the lesser of the two grids through the power supplies.
ps-redundant	This mode reserves the power of one supply in case any power supply fails. The power from the supply that can provide highest power is reserved. The available power is the sum of the remaining power supply units.
redundant	This mode combines power supply redundancy and input source redundancy, which means that the chassis has an extra power supply and each half of each power supply is connected to one electrical grid while the other half of each power supply is connected to the other electrical grid. The available power is the lesser of the available power for power supply mode and input source mode.

Fan Tray Redundancy

The Cisco Nexus 7000 series chassis contains two redundant system fan trays for I/O module cooling and two additional fan trays for switch fabric module cooling. Only one of each pair of fan trays is required to provide system cooling.

The fan speeds are variable and are automatically adjusted to one of 16 levels in order to optimize system cooling while minimizing overall system noise and power draw. A detected failure of a fan within a given fan tray will trigger an increase in the speed of the remaining fans to compensate for the failure. A detected removal of an entire fan tray, without replacement, will initiate a system shutdown after a three-minute warning period.

Starting with Cisco NX-OS Release 5.0(2a), the fan shutdown policy for the 10-slot chassis is as follows:

- If a system fan is removed: Earlier releases shut off the other fan in 3 minutes. The new policy is to increase the speed of the other fan based on the table mapping.
- If a fabric fan is removed: Earlier releases shut off the other fan in 3 minutes. The new policy is to increase the speed of the other fan to the maximum.

Caution In the case of a fan tray failure, In the Nexus 7009 or the Nexus 7018 devices, you must leave the failed unit in place to ensure proper airflow until a replacement is made available. The fan trays are hot swappable, but you must complete the removal and replacement within three minutes to avoid an automatic system shutdown.

Switch Fabric Redundancy

Cisco NX-OS provides switching fabric availability through redundant switch fabric module implementation. You can configure a single Nexus 7000 series with one to five switch fabric cards for capacity and redundancy. Each I/O module installed in the system automatically connects to and uses all functionally installed switch fabric modules. A failure of a switch fabric module triggers an automatic reallocation and balancing of traffic across the remaining active switch fabric modules. Replacing the failed fabric module reverses this process. After you insert the replacement fabric module and bring it online, traffic is again redistributed across all installed fabric modules and redundancy is restored.

Supervisor Module Redundancy

The Nexus 7000 series supports dual supervisor modules to provide 1+1 redundancy for the control and management plane. A dual supervisor configuration operates in an active or standby capacity in which only one of the supervisor modules is active at any given time, while the other acts as a standby backup. The state and configuration remain constantly synchronized between the two supervisor modules to provide statefu1 switchover in the event of a supervisor module failure.

Cisco NX-OS's Generic On-Line Diagnostics (GOLD) subsystem and additional monitoring processes on the supervisor trigger a stateful failover to the redundant supervisor when the processes detect unrecoverable critical failures, service restartability errors, kernel errors, or hardware failures.

If a supervisor-level unrecoverable failure occurs, the currently active, failed supervisor triggers a switchover. The standby supervisor becomes the new active supervisor and uses the synchronized state and configuration while the failed supervisor is reloaded. If the failed supervisor is able to reload and pass self-diagnostics, it initializes, becomes the new standby supervisor, and then synchronizes its operating state with the newly active unit.

Supervisor Restarts and Switchovers

Restarts on Single Supervisors

In a system with only one supervisor, when all HA policies have been unsuccessful in restarting a service, the supervisor restarts. The supervisor and all services reset and start with no prior state information.

Restarts on Dual Supervisors

When a supervisor-level failure occurs in a system with dual supervisors, the System Manager will perform a switchover rather than a restart to maintain stateful operation. In some cases, however, a switchover may not be possible at the time of the failure. For example, if the standby supervisor module is not in a stable standby state, a restart rather than a switchover is performed.

Switchovers on Dual Supervisors

A dual supervisor configuration allows nonstop forwarding (NSF) with stateful switchover (SSO) when a supervisor-level failure occurs. The two supervisors operate in an active/standby capacity in which only one of the supervisor modules is active at any given time, while the other acts as a standby backup. The two supervisors constantly synchronize the state and configuration in order to provide a seamless and stateful switchover of most services if the active supervisor module fails.

Switchover Characteristics

An HA switchover has the following characteristics:

- It is stateful (nondisruptive) because control traffic is not affected.
- It does not disrupt data traffic because the switching modules are not affected.
- Switching modules are not reset.
- It does not reload the Connectivity Management Processor (CMP).

  CMP is a Supervisor 1 only feature.

Switchover Mechanisms

Switchovers occur by one of the following two mechanisms:

- The active supervisor module fails and the standby supervisor module automatically takes over.
- You manually initiate a switchover from an active supervisor module to a standby supervisor module.

When a switchover process begins, another switchover process cannot be started on the same switch until a stable standby supervisor module is available.

Switchover Failures

If a switchover does not complete successfully within 28 seconds, the supervisors will reset. A reset prevents loops in the Layer 2 network if the network topology was changed during the switchover. For optimal performance of this recovery function, we recommend that you do not change the Spanning Tree Protocol (STP) default timers.

If three system-initiated switchovers occur within 20 minutes, all nonsupervisor modules will shut down to prevent switchover cycling. The supervisors remain operational to allow you to collect system logs before resetting the switch.

Manually Initiating a Switchover

To manually initiate a switchover from an active supervisor module to a standby supervisor module, use the **system switchover** command. After you run this command, you cannot start another switchover process on the same system until a stable standby supervisor module is available.

> **Note** If the standby supervisor module is not in a stable state (ha-standby), a manually-initiated switchover is not performed.

To ensure that an HA switchover is possible, use the **show system redundancy status** command or the **show module** command. If the command output displays the ha-standby state for the standby supervisor module, you can manually initiate a switchover.

Switchover Guidelines

Follow these guidelines when performing a switchover:

- When you manually initiate a switchover, it takes place immediately.
- A switchover can only be performed when two supervisor modules are functioning in the switch.
- The modules in the chassis must be functioning.

Verifying Switchover Possibilities

This section describes how to verify the status of the switch and the modules before a switchover.

- Use the **show system redundancy status** command to ensure that the system is ready to accept a switchover.
- Use the **show module** command to verify the status (and presence) of a module at any time. A sample output of the **show module**command follows:

```
switch# show module
Mod  Ports  Module-Type                         Model              Status
---  -----  ----------------------------------- ------------------ ----------
1    0      Supervisor module-1X                N7K-SUP1           active *
2    0      Supervisor module-1X                N7K-SUP1           ha-standby
3    32     1/10 Gbps Ethernet Module           N7K-D132XP-15      ok
4    48     1/10 Gbps Ethernet Module           N7K-F248XP-24      ok
5    48     10/100/1000 Mbps Ethernet XL Module N7K-M148GT-11L     ok
6    32     1/10 Gbps Ethernet Module           N7K-F132XP-15      ok
9    32     1/10 Gbps Ethernet Module           N7K-F132XP-15      ok

Mod  Sw              Hw
---  --------------  ------
1    6.0(1)          1.8
2    6.0(1)          1.1
3    6.0(1)          0.405
4    6.0(1)          0.500
5    6.0(1)          1.0
6    6.0(1)          0.617
9    6.0(1)          0.616


Mod  MAC-Address(es)                         Serial-Num
---  --------------------------------------  ----------
```

```
1    f0-25-72-ab-a3-f8 to f0-25-72-ab-a4-00  JAF1446BMRR
2    00-22-55-77-bc-48 to 00-22-55-77-bc-50  JAB122901WK
3    00-24-f7-1b-69-70 to 00-24-f7-1b-69-b4  JAF1321ARLQ
4    40-55-39-25-c8-00 to 40-55-39-25-c8-34  JAF1530AAAF
5    e8-b7-48-00-03-60 to e8-b7-48-00-03-94  JAF1513BPCH
6    f8-66-f2-02-a1-f8 to f8-66-f2-02-a2-3c  JAF1427DETN
9    a8-b1-d4-57-bc-bc to a8-b1-d4-57-bd-00  JAF1424CFMH

Mod  Online Diag Status
---  ------------------
1    Pass
2    Pass
3    Pass
4    Pass
5    Pass
6    Pass
9    Pass

Xbar Ports  Module-Type                         Model              Status
---  -----  ----------------------------------- ------------------ ----------
2    0      Fabric Module 2                     N7K-C7009-FAB-2    ok
4    0      Fabric Module 2                     N7K-C7009-FAB-2    ok
5    0      Fabric Module 2                     N7K-C7009-FAB-2    ok

Xbar Sw              Hw
---  --------------  ------
2    NA              0.201
4    NA              0.203
5    NA              0.201


Xbar MAC-Address(es)                         Serial-Num
---  --------------------------------------  ----------
2    NA                                      JAF1406ATRH
4    NA                                      JAF1422AHCP
5    NA                                      JAF1406ATRQ

* this terminal session
switch#
```

The Status column in the output should display an OK status for switching modules and an active or ha-standby status for supervisor modules.

- Use the **show boot auto-copy** command to verify the configuration of the auto-copy feature and if an auto-copy to the standby supervisor module is in progress. Sample outputs of the **show boot auto-copy** command are as follows:

```
switch# show boot auto-copy
Auto-copy feature is enabled
switch# show boot auto-copy list
No file currently being auto-copied
```

Replacing the Active Supervisor Module in a Dual Supervisor System

You can nondisruptively replace the active supervisor module in a dual supervisor system.

To replace the active supervisor module, follow these steps:

SUMMARY STEPS

1. switch # **system switchover**
2. switch# **out-of-service** *slot-number*
3. Remove the supervisor and insert the replacement.

DETAILED STEPS

	Command or Action	Purpose
Step 1	switch # **system switchover**	Initiates a manual switchover to the standby supervisor. **Note** Wait until the switchover completes and the standby supervisor becomes active.
Step 2	switch# **out-of-service** *slot-number*	Powers down the supervisor module you are replacing.
Step 3	Remove the supervisor and insert the replacement.	The new supervisor will automatically sync up the image and configuration from the currently active supervisor.

Replacing the Standby Supervisor Module in a Dual Supervisor System

You can nondisruptively replace standby supervisor module in a dual supervisor system.

To replace the standby supervisor module, follow these steps:

SUMMARY STEPS

1. switch# **out-of-service** *slot-number*
2. Remove the supervisor and insert the replacement.

DETAILED STEPS

	Command or Action	Purpose
Step 1	switch# **out-of-service** *slot-number*	Powers down the standby supervisor module.
Step 2	Remove the supervisor and insert the replacement.	The new supervisor will automatically sync up the image and configuration from the currently active supervisor.

Displaying HA Status Information

Use the **show system redundancy status** command to view the HA status of the system. The tables below explain the possible output values for the redundancy, supervisor, and internal states.

```
switch# show system redundancy status
Redundancy mode
---------------
      administrative:   HA
         operational:   HA
This supervisor (sup-1)
```

```
-----------------------
    Redundancy state:   Active
    Supervisor state:   Active
      Internal state:   Active with HA standby
Other supervisor (sup-2)
------------------------
    Redundancy state:   Standby
    Supervisor state:   HA standby
      Internal state:   HA standby
```

The following conditions identify when automatic synchronization is possible:

- If the internal state of one supervisor module is Active with HA standby and the other supervisor module is ha-standby, the system is operationally HA and can perform automatic synchronization.
- If the internal state of one of the supervisor modules is none, the system cannot perform automatic synchronization.

The Table below lists the possible values for the redundancy states.

Table 4: Redundancy States

State	Description
Not present	The supervisor module is not present or is not plugged into the chassis.
Initializing	The diagnostics have passed and the configuration is being downloaded.
Active	The active supervisor module and the switch are ready to be configured.
Standby	A switchover is possible.
Failed	The system detects a supervisor module failure on initialization and automatically attempts to power-cycle the module three times. After the third attempt, it continues to display a failed state.
Offline	The supervisor module is intentionally shut down for debugging purposes.
At BIOS	The system has established connection with the supervisor and the supervisor module is performing diagnostics.
Unknown	The system is in an invalid state. If it persists, call TAC.

This Table lists the possible values for the supervisor module states.

Table 5: Supervisor States

State	Description
Active	The active supervisor module in the switch is ready to be configured.
HA standby	A switchover is possible.
Offline	The system is intentionally shut down for debugging purposes.
Unknown	The system is in an invalid state and requires a support call to TAC.

This Table lists the possible values for the internal redundancy states.

Table 6: Internal States

State	Description
HA standby	The HA switchover mechanism in the standby supervisor module is enabled.
Active with no standby	A switchover is impossible.
Active with HA standby	The active supervisor module in the switch is ready to be configured. The standby supervisor module is in the ha-standby state.
Shutting down	The system is being shut down.
HA switchover in progress	The system is in the process of entering the active state.
Offline	The system is intentionally shut down for debugging purposes.
HA synchronization in progress	The standby supervisor module is in the process of synchronizing its state with the active supervisor modules.
Standby (failed)	The standby supervisor module is not functioning.
Active with failed standby	The active supervisor module and the second supervisor module are present but the second supervisor module is not functioning.
Other	The system is in a transient state. If it persists, call TAC.

Supervisor-to-Supervisor EOBC Link Redundancy

The Cisco Nexus 7000 Series switches provide two EOBC (Ethernet Out-of-Band) links for EOBC communication between the two supervisors. However, on switches running Cisco NX-OS Release 8.3(2) and older releases, only one EOBC link is utilized for EOBC communication. In case of an EOBC link failure, the active supervisor will try to reload the standby supervisor after a specific time. However, the standby supervisor will not come online leading to the loss of the high availability (HA) state.

Starting from Cisco NX-OS Release 8.4(1), the Supervisor-to-Supervisor EOBC Link Redundancy feature is introduced. This feature provides redundancy for EOBC communication between the active and standby supervisors by enabling the secondary redundant EOBC link in case the primary EOBC link fails and vice-versa. This feature is enabled by default.

Consider a scenario in which the Supervisor-to-Supervisor EOBC Link Redundancy feature is enabled and the primary EOBC link fails. This leads to the standby supervisor being reloaded by the active supervisor. While the standby supervisor is coming online, it will check if the primary supervisor-to-supervisor EOBC link is up by monitoring the EOBC heartbeat messages exchanged between the active and standby supervisors. In case the primary EOBC link is down, the standby supervisor will switch to the secondary redundant EOBC link. The standby supervisor will then come online. The active supervisor is also notified about the switch to the secondary redundant EOBC link and will use this link for any further EOBC communication.

Use the **show system internal redundancy status** command to display the currently operational EOBC link between the two supervisors.

```
switch# show system internal redundancy status
FSM state: RDN_DRV_ST_AC_SB
Current SUP states:
this : RDN_ST_AC (cmd reg = 62)
other: RDN_ST_SB (sts reg = 2630263)
Sysmgr status:
this : RDN_ST_AC
other: RDN_ST_SB
Slot: 3
Active Sup2Sup EOBC Link: Redundant
```

Syslog Generation

A syslog is generated by both the standby and active supervisors when the EOBC link is toggled to the secondary redundant EOBC link from the primary EOBC link or vice-versa.

```
Standby supervisor syslog:
System is coming up ... Please wait ...
System is coming up ... Please wait ...
2017 Dec 30 16:11:54 %$ VDC-1 %$ %KERN-2-SYSTEM_MSG: [ 365.924297] Not receiving EOBC
HB from active!!! Trying redundant sup2sup link - kernel //Indicates the toggle to
the
secondary redundant EOBC link
2017 Dec 30 16:11:54 %$ VDC-1 %$ %KERN-2-SYSTEM_MSG: [ 365.929695] sup2sup eobc using
redundant link - kernel
System is coming up ... Please wait ...
System is coming up ... Please wait ...
System is coming up ... Please wait ...
System is coming up ... Please wait ...
System is coming up ... Please wait ...
2017 Dec 30 16:12:23 switch %$ VDC-1 %$ %USBHSD-2-MOUNT: logflash: online
```

```
Active supervisor syslog:switch# 2017 Dec 30 16:11:53 switch %$ VDC-1 %$
%KERN-2-SYSTEM_MSG:
[23390.520159]sup2sup eobc using redundant link - kernel //Indicates that the
secondary
redundant EOBC link is in use after toggling from the primary EOBC link
```

Disabling Supervisor-to-Supervisor EOBC Link Redundancy

By default, the Supervisor-to-Supervisor EOBC Link Redundancy feature is enabled. To disable this feature, use the **no system eobc sup2sup link-redundancy** command in the global configuration mode.

Use the **show system internal redundancy status** command to verify that this feature is disabled.

```
switch# show system internal redundancy status
FSM state: RDN_DRV_ST_AC_SB
Current SUP states:
this : RDN_ST_AC (cmd reg = 2)
other: RDN_ST_SB (sts reg = 2630263)
Sysmgr status:
this : RDN_ST_AC
other: RDN_ST_SB
Slot: 3
Active Sup2Sup EOBC Link: PRIMARY (Sup2Sup EOBC Link redundancy is disabled)
```

If the secondary redundant EOBC link is active when you disable the Supervisor-to-Supervisor EOBC Link Redundancy feature, the standby supervisor is reloaded and it will come back online using the primary EOBC link. If the primary EOBC link is active when you disable the Supervisor-to-Supervisor EOBC Link Redundancy feature, disabling the Supervisor-to-Supervisor EOBC Link Redundancy feature will not have any impact on the supervisor HA state. However, a failure of the primary EOBC link will then not result in a switch to the secondary redundant EOBC link.

You can re-enable the Supervisor-to-Supervisor EOBC Link Redundancy feature by using the **system eobc sup2sup link-redundancy** command in the global configuration mode. If the standby supervisor is online when this command is used, there will be no impact on the current supervisor HA state. If the standby supervisor is not online when this command is used, the standby supervisor is reloaded and it will come back online using the currently working EOBC link.

Standby Supervisor Reload - Standby Supervisor in BIOS or Loader Prompt

When you reload the standby supervisor that is in BIOS or Loader prompt, there is no toggle from the primary EOBC link to the secondary redundant EOBC link or vice-versa. For example, if the secondary redundant EOBC link was active before reloading the standby supervisor, the same link will be active after the reload is complete and netboot will be initiated over the same link. In case you want to toggle from the secondary redundant EOBC link to the primary EOBC link, use the **system eobc sup2sup link-toggle** command in the global configuration mode of the active supervisor.

```
switch(config)# system eobc sup2sup link-toggle
sup2sup primary link became active. standby supervisor can be inserted or powered up now
```

Note Power off or remove the standby supervisor before using the **system eobc sup2sup link-toggle** command.

After using the **system eobc sup2sup link-toggle** command, power-on or insert the standby supervisor into the switch. The standby supervisor will then use the primary EOBC link while netbooting from the active supervisor.

Standby Supervisor Reload - Standby Supervisor not in BIOS or Loader Prompt

When you reload a standby supervisor that is not in BIOS or loader prompt, the standby supervisor comes back up with the system image and will then try to use the primary EOBC link. In case the primary EOBC link fails, it will toggle to the secondary redundant EOBC link. In case the secondary redundant EOBC link also fails, the standby supervisor will stop trying to toggle the EOBC links and it will fail to come online.

Toggling between the EOBC Links

After the standby supervisor is reloaded, by default, the BIOS will try to use the primary EOBC link if the standby supervisor attempts to perform a netboot from the BIOS. The netboot operation will fail if the primary EOBC link is down. In such a scenario, use the **system eobc sup2sup link-toggle** command to toggle to the secondary redundant EOBC link from the primary EOBC link. You can also use the same command to toggle to the primary EOBC link from the secondary redundant EOBC link.

```
switch(config)# system eobc sup2sup link-toggle
sup2sup redundant link became active. standby supervisor can be inserted or
powered up now
```

Note Power off or remove the standby supervisor before using the **system eobc sup2sup link-toggle** command.

After using the **system eobc sup2sup link-toggle** command, power-on or insert the standby supervisor into the switch. The standby supervisor will then use the secondary redundant EOBC link while netbooting from the active supervisor. In case the secondary redundant EOBC link was active before using the **system eobc sup2sup link-toggle** command, the standby supervisor will use the primary EOBC link while netbooting from the active supervisor.

System Switchover

Consider a scenario in which the previously active supervisor is coming up as the standby supervisor after reloading and the secondary redundant EOBC link is the active EOBC link. The standby supervisor will then try to use the primary EOBC link. If the primary EOBC link fails, it toggles to the secondary redundant EOBC link.

ISSU from Non-Supported Release to Supported Release

During the ISSU, if the standby supervisor is coming up with a release on which the Supervisor-to- Supervisor EOBC link redundancy feature is supported and the EOBC link fails, the standby supervisor attempts to toggle the EOBC link by notifying the active supervisor. However, the active supervisor does not respond as it is still running on a release image which does not support the Supervisor-to-Supervisor EOBC link redundancy feature. This results in the standby supervisor failing to come online and the ISSU is not completed.

Netbooting the Standby Supervisor during a Primary EOBC Link Failure

Consider a scenario in which the standby supervisor has to netboot over the EOBC from the active supervisor. By default, the standby supervisor uses the primary EOBC link to download the image. If the primary EOBC link is down, the netboot fails. You have to then toggle from the primary EOBC link to the secondary redundant EOBC link by using the **system eobc sup2sup link-toggle** command on the active supervisor.

```
switch(config)# system eobc sup2sup link-toggle
sup2sup redundant link became active. standby supervisor can be inserted or
powered up now
```

Note Power off or remove the standby supervisor before using the system eobc sup2sup link-toggle command.

After using the **system eobc sup2sup link-toggle** command, you can power-on or insert the standby supervisor. The standby supervisor will now use the secondary redundant EOBC link when it tries to netboot from the active supervisor.

Note The **system eobc sup2sup link-toggle** command will not toggle the EOBC link in case you have disabled the Supervisor-to-Supervisor EOBC Link Redundancy feature by using the **no system eobc sup2sup link-redundancy** command.

Enabling Supervisor-to-Supervisor EOBC Link Redundancy

The Supervisor-to-Supervisor EOBC Link Redundancy feature is enabled, by default, on Cisco NX-OS Release 8.4(1). Use this procedure to re-enable this feature in case you have disabled it.

Step 1 Enter global configuration mode:

switch# **configure terminal**

Step 2 Enable Supervisor-to-Supervisor EOBC Link Redundancy:

switch(config)# **system eobc sup2sup link-redundancy**

Step 3 Exit the global configuration mode:

switch(config)# **exit**

Step 4 (Optional) Display the currently operational EOBC link between the two supervisors:

switch# **show system internal redundancy status**

Running Configuration

This example shows a running configuration, followed by a verification command that displays the currently operational EOBC link between the two supervisors.

```
configure terminal
 system eobc sup2sup link-redundancy
 exit
.
.
.
switch# show system internal redundancy status
FSM state: RDN_DRV_ST_AC_SB
Current SUP states:
this : RDN_ST_AC (cmd reg = 62)
other: RDN_ST_SB (sts reg = 2630263)
Sysmgr status:
this : RDN_ST_AC
other: RDN_ST_SB
Slot: 3
Active Sup2Sup EOBC Link: Redundant
```

Disabling Supervisor-to-Supervisor EOBC Link Redundancy

Step 1 Enter global configuration mode:

switch# **configure terminal**

Step 2 Disable Supervisor-to-Supervisor EOBC Link Redundancy:

switch(config)# **no system eobc sup2sup link-redundancy**

Step 3 Exit the global configuration mode:

switch(config)# **exit**

Step 4 (Optional) Display the currently operational EOBC link between the two supervisors:

switch# **show system internal redundancy status**

Running Configuration

This example shows a running configuration, followed by a verification command that displays the currently operational EOBC link between the two supervisors.

```
configure terminal
 no system eobc sup2sup link-redundancy
 exit
.
.
.
switch# show system internal redundancy status
FSM state: RDN_DRV_ST_AC_SB
Current SUP states:
this : RDN_ST_AC (cmd reg = 2)
other: RDN_ST_SB (sts reg = 2630263)
Sysmgr status:
this : RDN_ST_AC
other: RDN_ST_SB
```

```
Slot: 3
Active Sup2Sup EOBC Link: PRIMARY (Sup2Sup EOBC Link redundancy is disabled)
```

Toggling between the EOBC Links

Step 1 Enter global configuration mode:

switch# **configure terminal**

Step 2 Toggle from the primary EOBC link to the secondary redundant EOBC link or vice-versa:

switch(config)# **system eobc sup2sup link-toggle**

Step 3 Exit the global configuration mode:

switch(config)# **exit**

Step 4 (Optional) Display the currently operational EOBC link between the two supervisors:

switch# **show system internal redundancy status**

Running Configuration

This example shows a running configuration for toggling from the secondary redundant EOBC link to the primary EOBC link, followed by a verification command that displays the currently operational EOBC link between the two supervisors.

```
configure terminal
 system eobc sup2sup link-toggle
 sup2sup primary link became active. standby supervisor can be inserted or
 powered up now
 exit
.
.
.
switch# show system internal redundancy status
FSM state: RDN_DRV_ST_AC_SB
Current SUP states:
this : RDN_ST_AC (cmd reg = 2)
other: RDN_ST_SB (sts reg = 2630263)
Sysmgr status:
this : RDN_ST_AC
other: RDN_ST_SB
Slot: 3
Active Sup2Sup EOBC Link: PRIMARY
```

This example shows a running configuration for toggling from the primary EOBC link to the secondary redundant EOBC link, followed by a verification command that displays the currently operational EOBC link between the two supervisors.

```
configure terminal
 system eobc sup2sup link-toggle
 sup2sup redundant link became active. standby supervisor can be inserted
 or powered up now
```

```
 exit
.
.
.
switch# show system internal redundancy status
FSM state: RDN_DRV_ST_AC_SB
Current SUP states:
this : RDN_ST_AC (cmd reg = 62)
other: RDN_ST_SB (sts reg = 2630263)
Sysmgr status:
this : RDN_ST_AC
other: RDN_ST_SB
Slot: 3
Active Sup2Sup EOBC Link: Redundant
```

VDC High Availability

The Cisco NX-OS software incorporates high availability (HA) features that minimize the impact on the data plane if the control plane fails or a switchover occurs. The different HA service levels provide data plane protection, including service restarts, stateful supervisor module switchovers, and in-service software upgrades (ISSUs). All of these high availability features support VDCs.

If unrecoverable errors occur in a VDC, the Cisco NX-OS software provides HA policies that you can specify for each VDC. These HA policies include the following:

- Bringdown—Puts the VDC in the failed state. To recover from the failed state, you must reload the physical device. This is the behavior for default VDC. For non-default VDC, there is no need to reload the physical device.
- Reset— Initiates a supervisor module switchover for a Cisco NX-OS device with two supervisor modules, or reloads a Cisco NX-OS device with one supervisor module.
- Restart—Deletes the VDC and recreates it by using the startup configuration.

For details about VDCs and HA, see the *Cisco Nexus 7000 Series NX-OS Virtual Device Context Configuration Guide*.

Related Documents

Related Topic	**Document Title**
Virtual device context (VDC)	*Cisco Nexus 7000 Series NX-OS Virtual Device Context Configuration Guide*
Redundant hardware	*Cisco Nexus 7000 Series Site Preparation Guide* and the *Cisco Nexus 7000 Series Hardware Installation and Reference Guide*
Power mode configuration and Cisco NX-OS fundamentals	*Cisco Nexus 7000 Series NX-OS Fundamentals Configuration Guide*
Nonstop forwarding (NSF)	*Cisco Nexus 7000 Series NX-OS Unicast Routing Configuration Guide*

Related Topic	Document Title
In-service software upgrades (ISSU)	*Cisco Nexus 7000 Series NX-OS Software Upgrade and Downgrade Guide*
GOLD, EEM, and Smart Call Home	*Cisco Nexus 7000 Series NX-OS System Management Configuration Guide*
Licensing	*Cisco Nexus 7000 Series NX-OS Licensing Guide*

Standards

Standards	Title
No new or modified standards are supported by this feature, and support for existing standards has not been modified by this feature.	—

MIBs

MIBs	MIBs Link
• CISCO-SYSTEM-EXT-MIB: ciscoHaGroup, cseSwCoresTable, cseHaRestartNotify, cseShutDownNotify, cseFailSwCoreNotify, cseFailSwCoreNotifyExtended • CISCO-PROCESS-MIB • CISCO-RF-MIB	To locate and download MIBs, go to the following URL: https://cfnng.cisco.com/mibs

RFCs

RFCs	Title
No RFCs are supported by this feature	—

Technical Assistance

Description	Link
Technical Assistance Center (TAC) home page, containing 30,000 pages of searchable technical content, including links to products, technologies, solutions, technical tips, and tools. Registered Cisco.com users can log in from this page to access even more content.	http://www.cisco.com/cisco/web/support/index.html

CHAPTER 6

ISSU and High Availability

This chapter describes the Cisco NX-OS in-service software upgrades (ISSU) and includes the following sections:

Information About ISSU

In a Nexus 7000 series chassis with dual supervisors, you can use the in-service software upgrade (ISSU) feature to upgrade the system software while the system continues to forward traffic. An ISSU uses the existing features of nonstop forwarding (NSF) with stateful switchover (SSO) to perform the software upgrade with no system downtime.

An ISSU is initiated through the command-line interface (CLI) by an administrator. When initiated, an ISSU updates (as needed) the following components on the system:

- Supervisor BIOS, kickstart image, and system image
- Module BIOS and image
- Connectivity Management Processor (CMP) BIOS and image

  CMP is a Supervisor 1 only feature.

In a redundant system with two supervisors, one of the supervisors is active while the other operates in the standby mode. During an ISSU, the new software is loaded onto the standby supervisor while the active supervisor continues to operate using the old software. As part of the upgrade, a switchover occurs between the active and standby supervisors, and the standby supervisor becomes active and begins running the new software. After the switchover, the new software is loaded onto the (formerly active) standby supervisor.

Virtualization Support

An ISSU-based upgrade is a system-wide upgrade that applies the same image and versions across the entire system, including all configured virtual device contexts (VDCs). VDCs are primarily a control-plane and user-interface virtualization and cannot run independent image versions per virtualized resource.

Note For complete information on VDCs, see the *Cisco Nexus 7000 Series NX-OS Virtual Device Context Configuration Guide*.

Guidelines and Limitations

An ISSU has the following limitations and restrictions:

- Do not change any configuration settings or network connections during the upgrade. Any changes in the network settings may cause a disruptive upgrade.
- In some cases, the software upgrades may be disruptive. These exception scenarios can occur under the following conditions:
    - A single supervisor system with kickstart or system image changes.
    - A dual-supervisor system with incompatible system software images.
- Configuration mode is blocked during the ISSU to prevent any changes.

For more information about compatible upgrades and downgrades, see *Cisco Nexus 7000 Series NX-OS Release Notes*.

How an ISSU Works

On a Nexus 7000 series with two supervisors, the ISSU process follows these steps:

1. Begins when the administrator uses the **install all** command
2. Verifies the location and integrity of the new software image files
3. Verifies the operational status and the current software versions of both supervisors and all switching modules to ensure that the system is capable of an ISSU
4. Loads the new software image to the standby supervisor and brings it up to the HA ready state
5. Forces a supervisor switchover
6. Loads the new software image to the (formerly active) standby supervisor and brings it up to the HA ready state
7. Performs a nondisruptive upgrade of each switching module, one at a time
8. Upgrades the Connectivity Management Processor (CMP)

   CMP is a Supervisor 1 only feature.

During the upgrade process, the system presents detailed status information on the console, requesting administrator confirmation at key steps.

ISSU and High Availability

This chapter describes the Cisco NX-OS in-service software upgrades (ISSU) and includes the following sections:

Determining ISSU Compatibility

An ISSU may be disruptive if you have configured features that are not supported on the new software image. To determine ISSU compatibility, use the **show incompatibility** system command.

This example shows how to determine ISSU compatibility:

```
switch# show incompatibility system bootflash:n7000-s1-dk9.4.1.4.bin
The following configurations on active are incompatible with  the system image
1) Service : vpc , Capability : CAP_FEATURE_VPC_ENABLED
Description : vPC feature is enabled
Capability requirement : STRICT
Disable command : Disable vPC using "no feature vpc"

2) Service : copp , Capability : CAP_FEATURE_COPP_DISTRIBUTED_POLICING
Description : Distributed policing for copp is enabled.
Capability requirement : STRICT
Disable command : Disable distributed policing using "no copp distributed-polici
ng enable”
```

Related Documents

Related Topic	Document Title
ISSU configuration	*Cisco Nexus 7000 Series NX-OS Fundamentals Configuration Guide*
Virtual device context (VDC)	*Cisco Nexus 7000 Series NX-OS Virtual Device Context Configuration Guide*
Licensing	*Cisco Nexus 7000 Series NX-OS Licensing Guide*

Standards

Standards	Title
No new or modified standards are supported by this feature, and support for existing standards has not been modified by this feature.	—

MIBs

MIBs	MIBs Link
• CISCO-SYSTEM-EXT-MIB: ciscoHaGroup, cseSwCoresTable, cseHaRestartNotify, cseShutDownNotify, cseFailSwCoreNotify, cseFailSwCoreNotifyExtended • CISCO-PROCESS-MIB • CISCO-RF-MIB	To locate and download MIBs, go to the following URL: https://cfnng.cisco.com/mibs

RFCs

RFCs	Title
No RFCs are supported by this feature	—

Technical Assistance

Description	Link
Technical Assistance Center (TAC) home page, containing 30,000 pages of searchable technical content, including links to products, technologies, solutions, technical tips, and tools. Registered Cisco.com users can log in from this page to access even more content.	http://www.cisco.com/cisco/web/support/index.html

INDEX