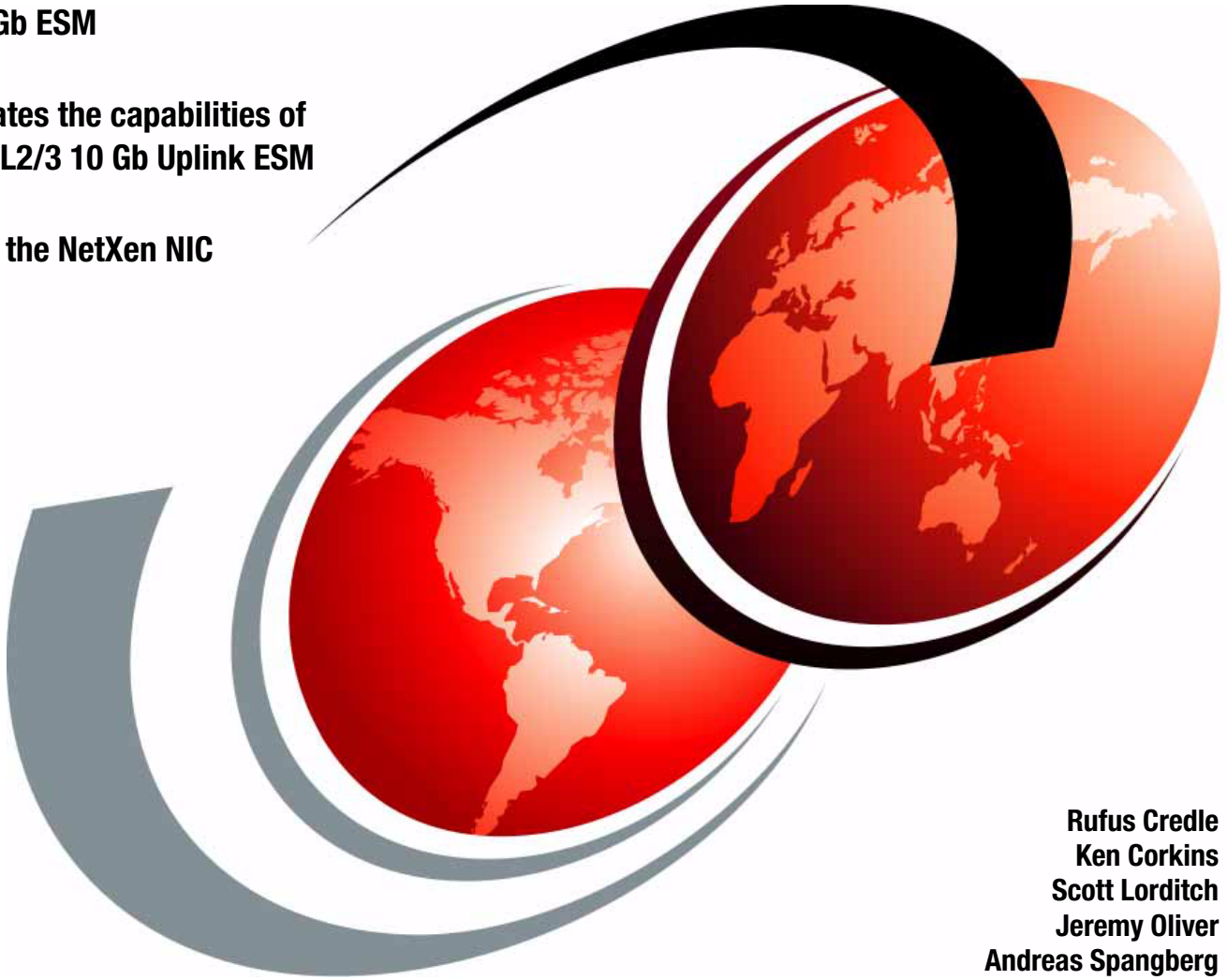
IBM

Networking with the Nortel 10 Gb Ethernet Switch Module for BladeCenter H

- Demonstrates the capabilities of the Nortel 10 Gb ESM
- Demonstrates the capabilities of the Nortel L2/3 10 Gb Uplink ESM
- Highlights the NetXen NIC



Rufus Credle
Ken Corkins
Scott Lorditch
Jeremy Oliver
Andreas Spangberg

Redpaper

ibm.com/redbooks

International Technical Support Organization

Networking with the Nortel 10 Gb Ethernet Switch Module for BladeCenter H

January 2008

REDP-4314-00

Note: Before using this information and the product it supports, read the information in “Notices” on page vii.

First Edition (January 2008)

This edition applies to Nortel 10 Gb Ethernet Switch Module for IBM BladeCenter.

Note: This paper is based on a pre-GA version of a product and may not apply when the product becomes generally available. We recommend that you consult the product documentation or follow-on versions of this paper for more current information.

Contents

Notices

This information was developed for products and services offered in the U.S.A.

IBM may not offer the products, services, or features discussed in this document in other countries. Consult your local IBM representative for information on the products and services currently available in your area. Any reference to an IBM product, program, or service is not intended to state or imply that only that IBM product, program, or service may be used. Any functionally equivalent product, program, or service that does not infringe any IBM intellectual property right may be used instead. However, it is the user's responsibility to evaluate and verify the operation of any non-IBM product, program, or service.

IBM may have patents or pending patent applications covering subject matter described in this document. The furnishing of this document does not give you any license to these patents. You can send license inquiries, in writing, to:
IBM Director of Licensing, IBM Corporation, North Castle Drive, Armonk, NY 10504-1785 U.S.A.

The following paragraph does not apply to the United Kingdom or any other country where such provisions are inconsistent with local law: INTERNATIONAL BUSINESS MACHINES CORPORATION PROVIDES THIS PUBLICATION "AS IS" WITHOUT WARRANTY OF ANY KIND, EITHER EXPRESS OR IMPLIED, INCLUDING, BUT NOT LIMITED TO, THE IMPLIED WARRANTIES OF NON-INFRINGEMENT, MERCHANTABILITY OR FITNESS FOR A PARTICULAR PURPOSE. Some states do not allow disclaimer of express or implied warranties in certain transactions, therefore, this statement may not apply to you.

This information could include technical inaccuracies or typographical errors. Changes are periodically made to the information herein; these changes will be incorporated in new editions of the publication. IBM may make improvements and/or changes in the product(s) and/or the program(s) described in this publication at any time without notice.

Any references in this information to non-IBM Web sites are provided for convenience only and do not in any manner serve as an endorsement of those Web sites. The materials at those Web sites are not part of the materials for this IBM product and use of those Web sites is at your own risk.

IBM may use or distribute any of the information you supply in any way it believes appropriate without incurring any obligation to you.

Information concerning non-IBM products was obtained from the suppliers of those products, their published announcements or other publicly available sources. IBM has not tested those products and cannot confirm the accuracy of performance, compatibility or any other claims related to non-IBM products. Questions on the capabilities of non-IBM products should be addressed to the suppliers of those products.

This information contains examples of data and reports used in daily business operations. To illustrate them as completely as possible, the examples include the names of individuals, companies, brands, and products. All of these names are fictitious and any similarity to the names and addresses used by an actual business enterprise is entirely coincidental.

COPYRIGHT LICENSE:

This information contains sample application programs in source language, which illustrate programming techniques on various operating platforms. You may copy, modify, and distribute these sample programs in any form without payment to IBM, for the purposes of developing, using, marketing or distributing application programs conforming to the application programming interface for the operating platform for which the sample programs are written. These examples have not been thoroughly tested under all conditions. IBM, therefore, cannot guarantee or imply reliability, serviceability, or function of these programs.

Trademarks

The following terms are trademarks of the International Business Machines Corporation in the United States, other countries, or both:

Redbooks (logo)  ®
eServer™
xSeries®
AIX®
BladeCenter®
IBM®
PowerPC®
Redbooks®
RETAIN®
System x™
Tivoli®
TotalStorage®
WebSphere®

The following terms are trademarks of other companies:

Network Appliance, NetApp, and the Network Appliance logo are trademarks or registered trademarks of Network Appliance, Inc. in the U.S. and other countries.

InfiniBand, and the InfiniBand design marks are trademarks and/or service marks of the InfiniBand Trade Association.

AMD, the AMD Arrow logo, and combinations thereof, are trademarks of Advanced Micro Devices, Inc.

Java, and all Java-based trademarks are trademarks of Sun Microsystems, Inc. in the United States, other countries, or both.

Convergence, Microsoft, Windows Server, Windows, and the Windows logo are trademarks of Microsoft Corporation in the United States, other countries, or both.

Intel, Intel logo, Intel Inside logo, and Intel Centrino logo are trademarks or registered trademarks of Intel Corporation or its subsidiaries in the United States, other countries, or both.

UNIX is a registered trademark of The Open Group in the United States and other countries.

Linux is a trademark of Linus Torvalds in the United States, other countries, or both.

Other company, product, or service names may be trademarks or service marks of others.

NetXen, the NetXen logo, Intelligent NIC, Linux Sockets Acceleration and Selective Acceleration are trademarks of NetXen. All other company names and trademarks are the property of their respective holders.

Preface

This Redpaper discusses the physical layout of the 10 Gb switch and the installation in HSS slots 7 or 9 in BladeCenter® H chassis. In addition, we discuss the physical layout of the NetXen HBA card and its installation in the HS21/LS21 and connectivity to an IBM® TotalStorage® device.

This Redpaper provides detailed instructions on how to install and configure NetXen drivers on the Windows® 2K3 and Linux® (RHEL and SUSE) platforms.

This Redpaper discusses:

- NIC Teaming and Bonding
- The 10 Gb switch configuration
  - Cisco integration, including Link Aggregation, multiple VLANs (802.1q), L3 routing with OSPF, Spanning Tree modes
  - HA features: Trunk failover, VRRP
  - Filters/access lists, including use of QoS; can include QoS example for iSCSI or other
- 10 Gb iSCSI
  - with N-Series/NetApp® storage with 10 Gb NIC
  - Best performance with TCP Chimney or LSA

The team that wrote this paper

This paper was produced by a team of specialists from around the world working at the International Technical Support Organization, Raleigh Center.

From left to right, Scott Lorditch, Jeremy Oliver, Ken Corkins, Andreas Spangberg, and Rufus Credle

Rufus Credle is a Certified Consulting IT Specialist at the ITSO, Raleigh Center. In his role as project leader, he conducts residencies and develops IBM Redbooks® about network operating systems, ERP solutions, voice technology, high availability and clustering solutions, Web application servers, pervasive computing, IBM and OEM e-business applications, IBM System x™, xSeries® and IBM BladeCenter. Rufus' various positions during his IBM career have included assignments in administration and asset management, systems engineering, sales and marketing, and IT services. He holds a BS degree in business management from Saint Augustine's College. Rufus has been employed at IBM for 27 years.

Ken Corkins is a member of the System x Advanced Technical Support Group in Dallas, Texas. He specializes in BladeCenter and networking. His responsibilities include assisting BladeCenter pre-sale customers in the Americas with proof of concepts and pilot testing. He organized the Nortel Champions program for North America in Dallas. He has 12 years of experience in the IT industry and holds the Cisco CCNP and CCDP certifications. Ken started with IBM in October 2000 as an IT Specialist (Networking) with IBM Global Services.

Scott Lorditch is a Sales Network Architect for Blade Network Technologies, Inc. (formerly the Blade Switching Server business unit of Nortel Networks). He develops designs and proposals for customers and potential customers of Blade Network Technologies, Inc. GbESM products for the IBM BladeCenter, including overall network architecture assessments. He has also developed several training and lab sessions for IBM technical and sales personnel and has provided field feedback to the product team. His background before working for Blade Network Technologies, Inc. includes almost 20 years working on networking including electronic securities transfer projects for a major bank based in New York City, as a Senior Network Architect for a multi-national soft drink company, and as Product Manager for managed hosting services for a large telecommunications provider. He holds a BS in Operations Research with specialization in Computer Science from Cornell University.

Jeremy Oliver is a licensed Professional Engineer of North Carolina and works in IBM Systems and Technology Group, System x. He has worked at IBM Research Triangle Park, North Carolina for nine years. His educational background includes a BS degree in Electrical Engineering from McNeese State University, Lake Charles, Louisiana. He also holds a MS degree in Electrical Engineering from North Carolina State University, Raleigh, with research in PHY technologies of computer networking. His area of expertise is developing experiments for testing new BladeCenter technologies, focusing on switch I/O modules. Jeremy was a co-author of REDP-3586-00 on Nortel L2/3 ESM, published in 2005.

Andreas Spangberg is a Staff Engineer at the ISS Wintel Technical Support Centre in IBM UK in Greenock. He works in the EMEA System-X and BladeCenter Level 2 department as BladeCenter and Blade Network Option PFE and is Technical Team Lead of the Level 2 Blade team. Andreas has 12 years experience in IBM Technical Support. He specializes in BladeCenter and Networking and holds the CISCO CCNA and CCNP certifications. He organized the Nortel Champions program for EMEA in Greenock.

Thanks to the following people for their contributions to this project by providing hardware, advice, and support:

Tamikia Barrow, Carolyn Briscoe, Margaret Ticknor
International Technical Support Organization, Raleigh Center

Ishan Sehgal, IBM Systems &Technology Group, System x; BladeCenter Marketing Manager, Networking; Offering Management Marketing Manager
IBM RTP

Cynthia Knight, IBM Systems &Technology Group, Development, Hardware Design
IBM RTP

Rajesh Vaidyanath, NetXen, Inc.
NetXen Inc.

Alexander Rossetto, BladeCenter Product Engineer
IBM RTP

William Champion, BladeCenter Infrastructure Solution Center Team
IBM RTP

Mark Welch, BladeCenter Development
IBM RTP

Eva Ho, IBM Systems &Technology Group, Development
IBM RTP

Khalid Ansari, IBM Sales & Distribution, TSS
IBM RTP

Mats Wahlstrom,
NetApp RTP

Become a published author

Join us for a two- to six-week residency program! Help write a book dealing with specific products or solutions, while getting hands-on experience with leading-edge technologies. You will have the opportunity to team with IBM technical professionals, Business Partners, and Clients.

Your efforts will help increase product acceptance and customer satisfaction. As a bonus, you will develop a network of contacts in IBM development labs, and increase your productivity and marketability.

Find out more about the residency program, browse the residency index, and apply online at:

ibm.com/redbooks/residencies.html

Comments welcome

Your comments are important to us!

We want our papers to be as helpful as possible. Send us your comments about this paper or other IBM Redbooks in one of the following ways:

- Use the online **Contact us** review form found at:

  ibm.com/redbooks

- Send your comments in an e-mail to:

  redbooks@us.ibm.com

- Mail your comments to:

  IBM Corporation, International Technical Support Organization
  Dept. HYTD Mail Station P099
  2455 South Road
  Poughkeepsie, NY 12601-5400

1

Executive summary

IBM and Nortel Networks are committed to collaborating on the design and development of server and networking technology to address customer requirements by establishing a joint development center. The Nortel 10 Gb Ethernet Switch Module for IBM BladeCenter represents a new height in this alliance.

The BladeCenter switch module offers BladeCenter customers Nortel's latest 10 Gb Gigabit Ethernet switching technology which is integrated into the BladeCenter H high speed chassis. It further enhances the BladeCenter value proposition by interfacing the NetXen 10 Gb Ethernet Expansion Card for IBM eServer BladeCenter option to a customer's existing data network using six external Fiber 10 Gb XFP interconnect options (SR or LR). When installed in the BladeCenter H chassis, the Nortel 10 Gb Ethernet Switch Module provides both full L2 switching and L3 routing capabilities and significant added value not found in commodity switching solutions. This value includes:

- 10 Gb Ethernet - 802.3ae
- VLAN tagging - 802.1Q
- Link Aggregation and LACP - 802.3ad and 802.3-2002
- Spanning Tree - 802.1D, 802.1w, 802.1s
- Routing Information Protocol - RFC1058 and RFC2453
- Open Shortest Path First (OSPF) - RFC1257, RFC2328
- Routing Information Protocol (RIP) Versions 1 and 2
- Border Gateway Protocol Version 4 (BGPv4)
- Virtual Router Redundancy Protocol (VRRP) - RFC 3768
- Support of iSCSI and TCP/IP offload protocols

Each Nortel 10 Gb Ethernet Switch Module provides 10 Gigabit per second Ethernet (10 GbE) connectivity to each of the 14 blade slots and six 10 GbE uplink interfaces external to the IBM BladeCenter. The customer can install one or two Nortel 10 Gb Ethernet Switch Modules in one BladeCenter H chassis. With two Nortel 10 Gb Ethernet Switch Modules installed, you can obtain 12 10 GbE uplink interfaces as well as 28 10 GbE internal switching interfaces. The flexibility of the Nortel 10 Gb Ethernet Switch Module allows you to address a variety of performance and redundancy needs.

The Nortel and IBM agreement to form a joint development center equips Nortel as it becomes an on demand company that can generate customized products for its network equipment marketplace. This ensures that your needs of high availability, scalability, security, and manageability are addressed. Combined with the integration of IBM Tivoli®, Nortel, and Cisco management products, these architectures bring higher value solutions with lower operational expense.

The Nortel 10 Gb Ethernet High Speed Switch Module is an integral part of these solutions. With the Nortel 10 Gb Ethernet Switch Module, you have the investment protection and price performance of a solution behind which the world's leading server and networking companies stand.

1.1 Nortel L2/3 10 Gb Switch Module value proposition

This section discusses the value of using the Nortel L2/3 10 Gb Ethernet Switch Module for your IBM BladeCenter:

Product strength

The product provides strengths such as:

- Provides full interoperability into existing Nortel and Cisco networks
- Integrates Nortel networking capabilities to reduce data center complexity and increases networking manageability and availability
- Leverages the leadership capabilities of the members of Blade.org to provide the most technological choices

Leadership features and function

The leadership features and function include:

- The IBM BladeCenter H/HT delivers with the 10 Gb Ethernet Switch Module as much as 200 Gbps of Ethernet bandwidth per switch with full Layer 2 switching, Layer 3 switching (routing), and Layer 4 filtering and related services.
- The switch module runs Alteon Operating System and appears as any other product from Nortel's Alteon product line to the data center's network management tools. In addition, the 10 Gb Uplink Switch Module has the ability to use an industry standard Command Line Interface (isCLI) which will be familiar to most network technical personnel.

Competitive advantage

The product delivers a competitive advantage by delivering:

- Full integration of 10 Gb Ethernet switching, reducing infrastructure complexity
- Six external 10 Gb fiber ports
- Price leadership in the 10 Gb Ethernet switching space

1.2 Nortel L2/3 10 Gb Uplink Switch Module value proposition

This section discusses the value of using the Nortel L2/3 10 Gb Uplink Ethernet Switch Module for your IBM BladeCenter H or HT.

Product strength

The product provides strengths such as:

- Provides full interoperability into existing Nortel and Cisco networks
- Integrates Nortel networking capabilities to reduce data center complexity and increases networking manageability and availability
- Leverages the leadership capabilities of the members of Blade.org to provide the most technological choices

Leadership features and function

The leadership features and function include:

- The IBM BladeCenter H/HT delivers with the 10 Gb Uplink Ethernet Switch Module as much as 45 Gbps of Ethernet bandwidth with full Layer 2 switching and Layer 3 switching (routing) functionality as well as Layer 4 filtering and related services.
- The switch module runs the Alteon Operating System and appears as any other product from Nortel's Alteon product line to the data center's network management tools. In addition, the 10 Gb Switch Module has the ability to use an industry standard Command Line Interface (isCLI) which will be familiar to most network technical personnel.

Competitive advantage

The product delivers a competitive advantage by delivering:

- Full integration of 10 Gb Ethernet switching, reducing infrastructure complexity
- Three external 10 Gb ports (two CX4 ports and one fiber port)
- Price leadership in the 10 Gb Ethernet switching space

Items that must be mentioned:

- BNT background
- Blade.org
- Discussion of applicability of HBC clusters
- Discussion regarding iSCSI

2

BladeCenter H and HT chassis

This chapter discusses the differences between the BladeCenter H and HT models. This chapter also covers Ethernet connectivity in the chassis, including NIC bonding.

2.1 BladeCenter H

Figure 2-1 IBM BladeCenter H Chassis front view

The IBM BladeCenter H (as shown in Figure 2-1) is a high-performance 9U chassis, designed for compute-intensive environments, such as Earth/Life Sciences, commercial analytics and next-generation network applications. It provides:

- Reduced single points of failure — Many major components (either standard or optionally) are hot-swappable or redundant. Servers and modules can be configured for automatic failover to backups.
- Forward and backward compatibility — Every blade, switch, and passthru module released by IBM for the original BladeCenter chassis since 2002 is supported in the BladeCenter H chassis.
- High-speed redundant midplane connections — Based on 4X InfiniBand®, the midplane supports up to 40 Gb bandwidth and provides four 10 Gb data channels to each blade. By giving each blade two physical connections to the midplane that connects all blades and modules together internally, a failure of one connector alone cannot bring down the server.
- Fourteen 30 mm blade slots — These hot-swap slots are capable of supporting any combination of 14 HS20/HS21/HS21 XM (Xeon), LS20/LS21 (Opteron), and JS20/JS21 (PowerPC®) blade servers, or seven double-wide (60 mm) HS40 or LS41 blade servers, or a mixture of 30 mm and 60 mm blades. It also supports multiple optional 30 mm Expansion Units in combination with the blade servers, using the same blade slots. Up to four chassis can be installed in an industry-standard 42U rack, for a total of up to 56 30 mm blade servers per rack.
- Up to ten module bays for communication and I/O switches or bridges — The modules interface with all of the blade servers in the chassis and alleviate the need for external switches or expensive, cumbersome, and error-prone cabling. All connections are done internally through the midplane. Two module slots are reserved for hot-swap/redundant Gigabit Ethernet switch modules. Two slots support either high-speed bridge modules or traditional Gigabit Ethernet, Myrinet, Fibre Channel, InfiniBand, and other switch modules. Two slots are dedicated for bridge modules. Four additional slots are dedicated for hot-swap/redundant high-speed switch modules. All modules, when installed in pairs, offer load balancing and failover support. Figure 2-2 on page 7 details the module bay locations.
- Integrated switch and bridge modules mean that no additional rack U space is required.

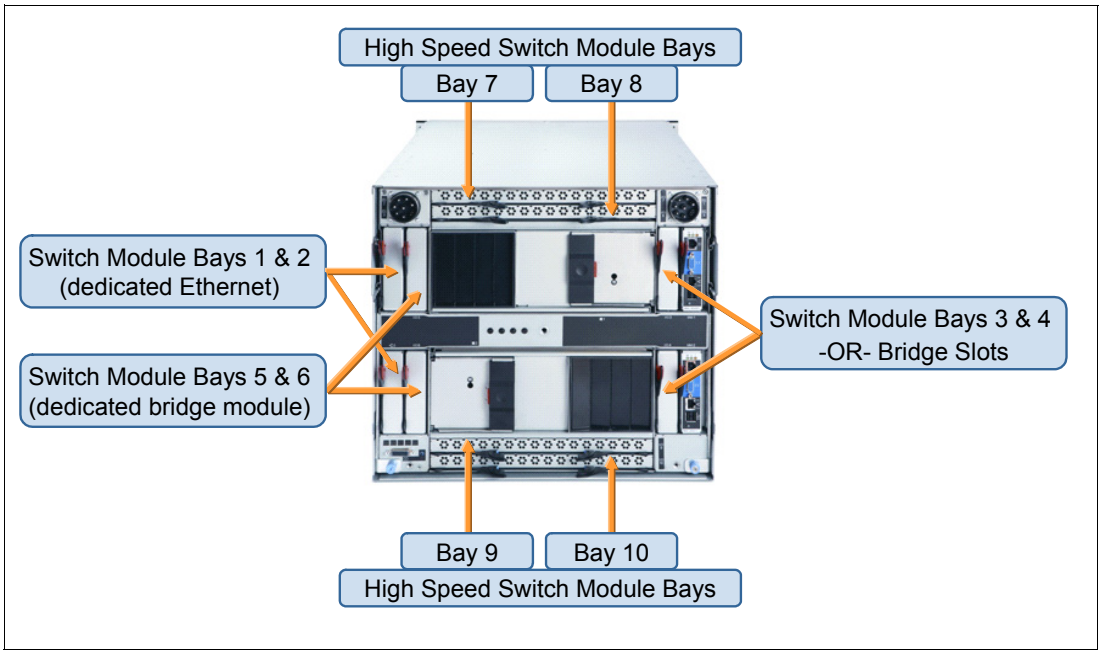


Figure 2-2 IBM BladeCenter H chassis switch module identification

- Two module bays for Advanced Management Modules — The management modules provide advanced systems management and KVM capabilities for not only the chassis itself, but for all of the blades and other modules installed in the chassis. The Advanced Management Module provides capabilities similar to the IBM Remote Supervisor Adapter II SlimLine used in stand-alone System x rack and tower servers. Features include concurrent KVM (cKVM), an external Serial over LAN connection, industry-standard management interfaces (SMASH/CLP/CIM/HPI), USB virtualization, network failover and backward compatibility with the original Management Module, among others. The features of the module can be accessed either locally or remotely across a network. One module comes standard. A second module can be added for hot-swap/redundancy and failover. The module uses USB ports for keyboard and mouse.
- Two module bays for Blower Modules — Two hot-swap/redundant blower modules come standard with the chassis. They are capable of providing efficient cooling for up to 14 blades. These modules replace the need for each blade and switch to contain its own fans. The blowers are more energy efficient than dozens or hundreds of smaller fans would be, and they offer fewer points of potential failure. BladeCenter H also includes up to four additional hot-swap/redundant fan packs to cool the power supplies and high-speed switch modules.
- Four bays for Power Modules — BladeCenter H ships with two 2900W high-efficiency hotswap/redundant power modules (upgradeable to four), capable of handling the power needs of the entire chassis, including future higher-wattage processors. Each power module includes a customer-replaceable hot-swap/redundant fan pack (three fans) for additional cooling capability.
- A hot-swappable Media Tray containing a DVD-ROM drive, two USB 2.0 ports, and a light path diagnostic panel — The media tray is shared by all the blades in the server. This reduces unnecessary parts (and reduces the number of parts than can fail). In the event of a failure of the Media Tray, the tray can be swapped for another. While the tray is offline, the servers in the chassis can remotely access the Media Tray in another chassis. The light path diagnostic panel contains LEDs that indicate chassis status.

2.2 BladeCenter HT

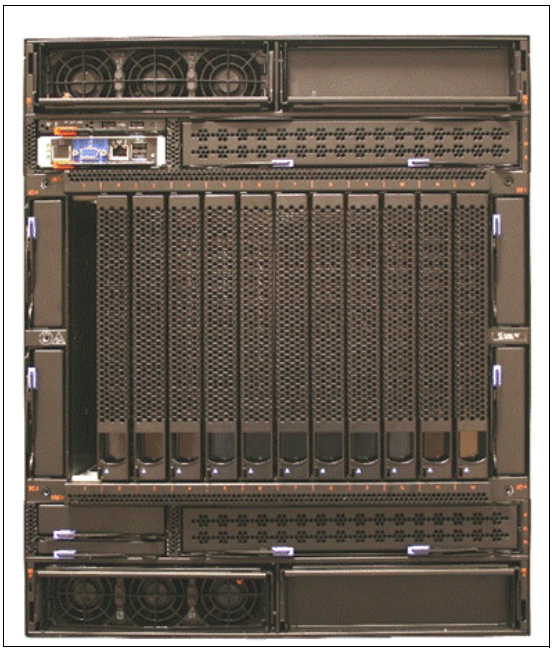


Figure 2-3 IBM BladeCenter HT chassis front view

BladeCenter HT (Figure 2-3) is a carrier grade, rugged 12U chassis designed for challenging central office and networking environments. It provides:

- NEBS Level 3/ETSI-tested — Designed for the Network Equipment Provider (NEP) and Service Provider (SP) environments. Also ideal for government, military, aerospace, industrial automation and robotics, medical imaging, and finance. Certified testing by Underwriters Laboratories of the BladeCenter HT chassis is in progress; when complete, it will be covered under a UL-certified NEBS Level 3/ETSI test report.
- Designed for Carrier-Grade Linux — Several distributions are supported, include SUSE and Red Hat.
- Reduced single points of failure — Many major components (either standard or optionally) are hot-swappable and/or redundant. Servers and modules can be configured for automatic failover to backups.
- Backward compatibility — Every blade, switch, and passthru module released by IBM for the original BladeCenter chassis since 2002 is supported in the BladeCenter HT chassis.
- High-speed redundant midplane connections — Based on 4X InfiniBand, the midplane supports up to 40 Gb bandwidth and provides four 10 Gb data channels to each blade. By giving each blade two physical connections to the midplane that connects all blades and modules together internally, a failure of one connector alone cannot bring down the server.
- Twelve 30 mm blade slots — These hot-swap slots are capable of supporting any combination of 12 HS20/HS21/HS21 XM (Xeon), LS20/LS21 (Opteron), and JS20/JS21 (PowerPC) blade servers, or 6 double-wide (60 mm) LS41 blade servers, or a mixture of 30 mm and 60 mm blades. It also supports multiple optional 30 mm Expansion Units in combination with the blade servers, using the same blade slots. Up to three chassis can be installed in an industry-standard 42U rack, for a total of up to 36 30 mm blade servers per rack.

- Two module bays for Advanced Management Modules — The management modules provide advanced systems management and KVM capabilities for not only the chassis itself, but for all of the blades and other modules installed in the chassis.
- Four bays for Fan Modules — All four hot-swap/redundant fan modules come standard with the chassis. These modules replace the need for each blade to contain its own fans. The high availability modules are more energy efficient than dozens or hundreds of smaller fans would be, and there are fewer points of potential failure.
- Four bays for Power Modules — BladeCenter HT ships with two high-efficiency hotswap/redundant DC or AC (model-specific) power modules (upgradeable to four), capable of handling the power needs of up to six blade servers. Two additional power modules are required when more than six blades or high-speed switches are installed.
- Two hot-swappable Media Trays each contain two external USB 2.0 ports, a light path diagnostic panel, and support a 1 Gb/4 Gb compact flash (CF) option — The media tray is shared by all the blades in the server. This reduces unnecessary parts (and reduces the number of parts than can fail). In the event of a failure of the Media Tray the tray can be swapped for another. While the tray is offline, the servers in the chassis can remotely access the Media Tray in another chassis. The light path diagnostic panel contains LEDs that indicate chassis status. One media tray comes standard (without compact flash); an optional second one provides redundancy. The CF option can act as a boot device, eliminating the need for HDDs in the blades.
- Redundant midplane connections — Each chassis contains a midplane that connects all blades and modules together internally. The midplane provides two physical connections to each blade; therefore, a failure of one connector alone cannot bring down the server.

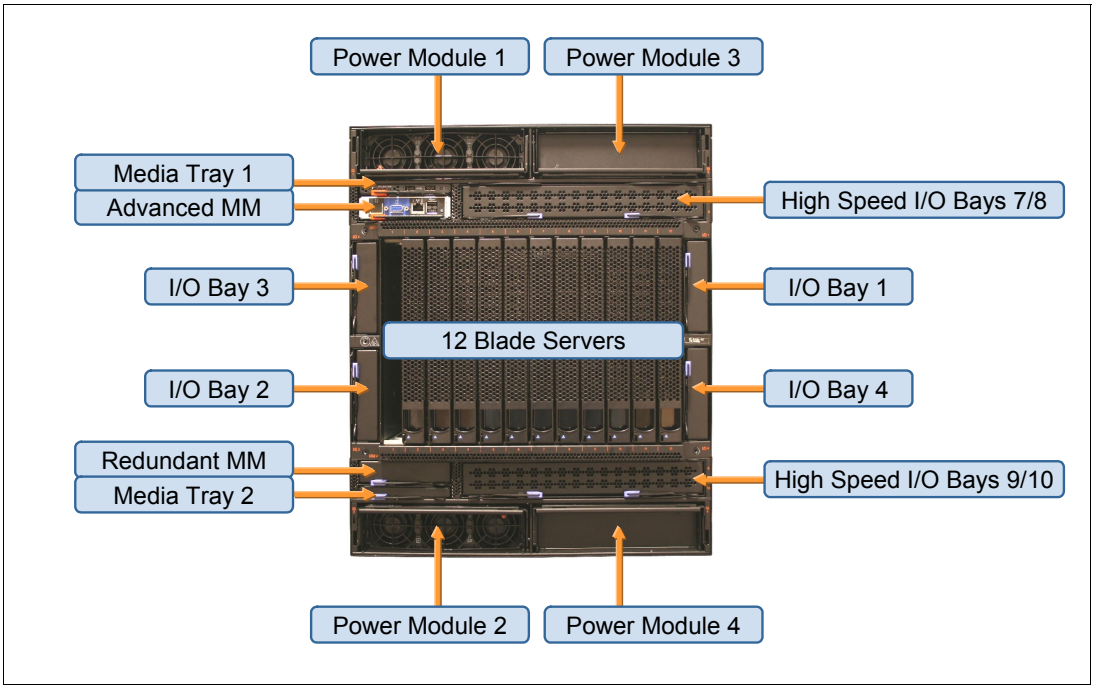


Figure 2-4 BladeCenter HT chassis module locations

- Up to eight module bays for communication and I/O switches or bridges — The modules interface with all of the blade servers in the chassis and alleviate the need for external switches or expensive, cumbersome cabling. All connections are done internally through the midplane. Two module slots are reserved for hot-swap/redundant Gigabit Ethernet switch modules. Two slots support either high-speed bridge modules or traditional Gigabit Ethernet, Myrinet, Fibre Channel, InfiniBand, and other switch modules. Four additional

slots are dedicated for hotswap/redundant high-speed switch modules. All modules, when installed in pairs, offer load balancing and failover support.

- Bridge Modules — Dedicated bridge module bays do not exist in BCHT. I/O bays 3 and 4 are similar to BCH, but bridge bays 5 and 6 do not exist. BCHT I/O bays 1 and 2 do provide bridge module support but no hybrid bridge/Gb switch modules are planned at this time.
- Interposers — AMMs and I/O (switch) modules are all located in the front of the chassis with the blades. Since they plug into a common midplane with the blade servers, they require interposers in each populated I/O bay to make up the difference in depth between switch enclosures and blade server enclosures. These interposers are entirely passive, and extend the midplane out to the rear connectors of the I/O modules. See Figure 2-5 for a picture.

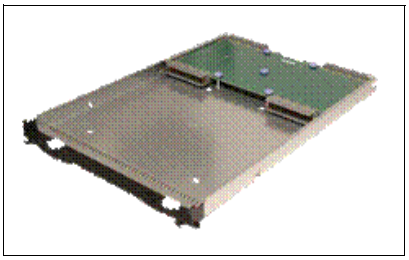


Figure 2-5 High-Speed switch tray with interposer

- Interswitch Links — Because BCHT has two unimplemented blade ports (13 and 14), the BCHT midplane provides interconnects between switch ports 13 and 14 for the following matched pairs of switch modules: 1-2, 3-4, 7-9, 8-10. These must be enabled by specifically-designed switch interposers that complete the link between the switch module ports 13 and 14 and the midplane paths. If matching interposers do not both have the ISL support, then the switch module ports will not be connected. Interswitch links between matched pairs of switch modules allow additional functionality such as link aggregation.

2.3 Blade connections

A single IBM BladeCenter HS21, JS21, or LS21 blade is provided 20 lanes to the BladeCenter H midplane, as follows:

- Four single-lane connections to four standard switch module bays
- Four 4-lane connections to four high-speed switch module bays

With the newest blade servers, HS21 (8853), LS21 (7971), LS41 (7972), and JS21 (8844), new daughter card form factors have been introduced: the high-speed form factor, the Combo Form Factor Horizontal (CFF-h), and Combo Form Factor Vertical (CFF-v). High-speed form factor and CFF-h cards connect to the high-speed switch bays, 7-10.

2.4 Ethernet switch modules and network interfaces in the IBM BladeCenter H

Figure 2-6 is a depiction of a blade server in a BladeCenter chassis. The example blade shows two onboard NICs (NIC0 and NIC1). NIC0 is connected through the BladeCenter midplane to switch bay one. NIC1 is connected to switch bay two. These connections are hard-wired in the chassis mid-plane and cannot be changed. Switch bays one and two in the BladeCenter H and BladeCenter HT are dedicated Ethernet switch bays, meaning, only Ethernet capable devices can be installed in these two switch bays.

Some blade servers like the LS41 (MT7972) are 60 mm wide and occupy two server slots in the chassis. The LS41 (and other double-wide blades) have an additional pair of Ethernet interfaces on the second module of the blade. These additional NICs (depicted as NIC2 and NIC4) are also hard wired to switch bays one and two, respectively.

Switch bays three and four are optional switch bays. These switch bays can be used for Ethernet connectivity if desired by installing an optional Ethernet daughter card (CFF-v) on the blade server. Interface 0 of the daughter card (INTF0) is connected to switch bay three and interface 1 (INTF1) is connected to switch bay four. Again, these connections cannot be changed. Appropriate Ethernet switch modules must also be installed in switch bays three and four to support Ethernet connectivity to the CFF-v Ethernet card.

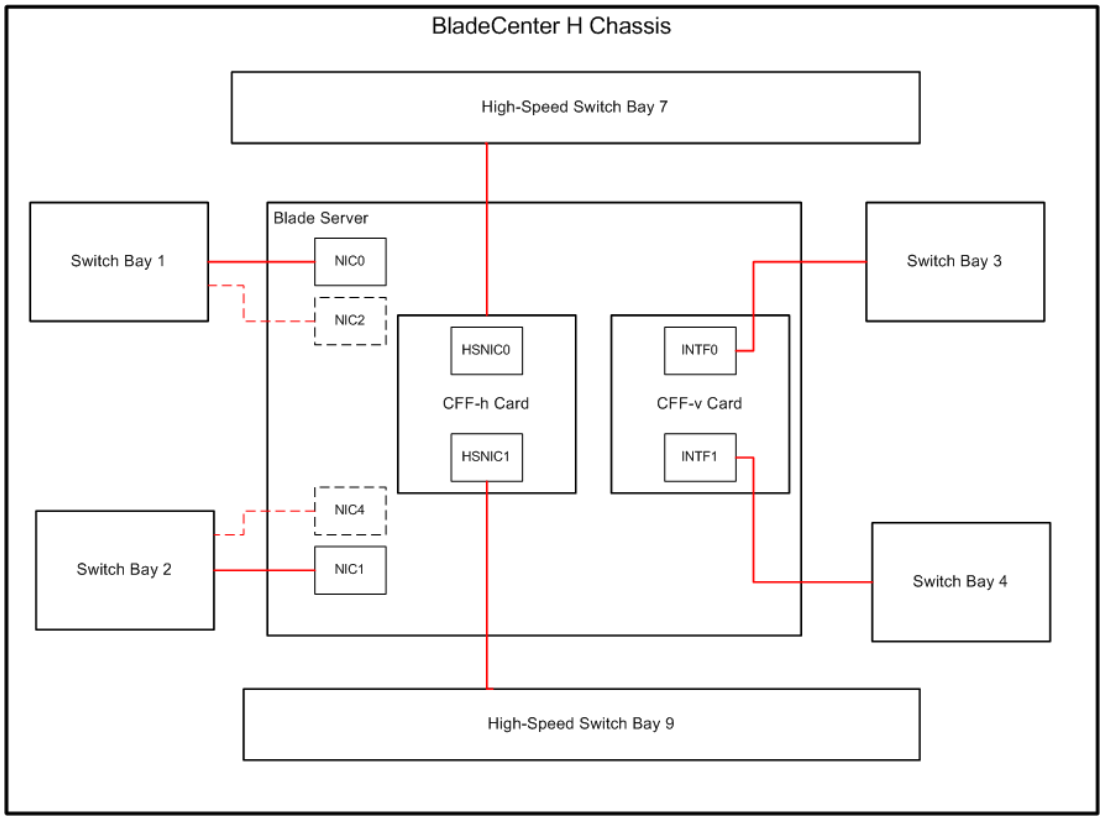


Figure 2-6 BladeCenter H Ethernet switch module and blade server NIC connections

Switch bays seven and nine are optional high-speed switch bays. These switch bays are used for Ethernet connectivity if desired by installing the Nortel 10 Gb switch module and an optional Ethernet daughter card (CFF-h) on the blade server. Interface 0 of the daughter card

(HSNIC0) is connected to switch bay seven and interface 1 (HSNIC1) is connected to switch bay nine. Again, these connections cannot be changed.

The BladeCenter HT chassis has additional ISL links between the switch bays, as depicted in Figure 2-7. These interconnects do not change how the server's NICs interface to the switches, but allow for internal switch interconnection. The ISL interposers are optional for switch bays one and two, and three and four, but are not optional for switch bays seven and nine. Care must be taken when initially configuring switch modules where ISL interposers are involved. Network loops can be created inside the chassis if the spanning-tree protocol is not configured on ports 13 and 14 of the switch modules.

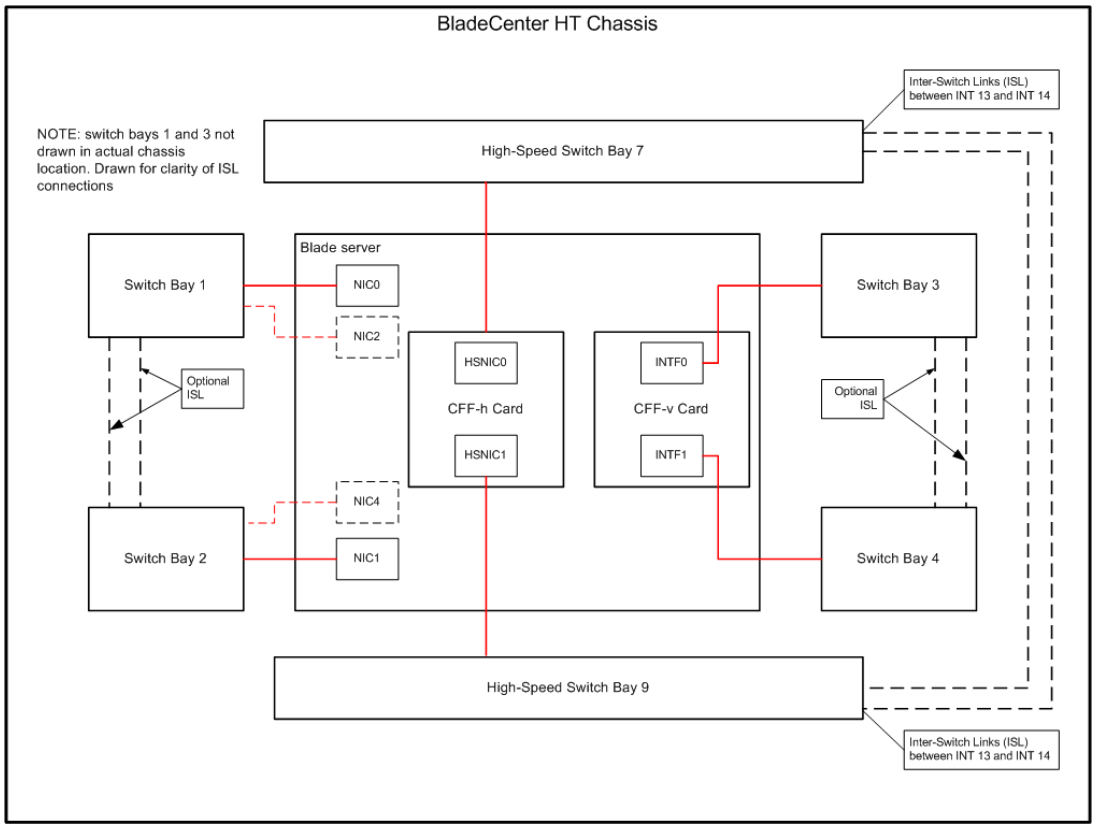


Figure 2-7 BladeCenter HT Ethernet switch module and blade server NIC connections

2.5 NIC bonding

NIC bonding is the ability to combine two (or more) Ethernet interfaces on a server into a single logical network interface. Bonding is generally a term used in Linux, whereas the term teaming is used in Windows. In a bonding configuration, a new logical interface is created using an Intermediate (IM) driver. NIC bonding can be used to aggregate the bandwidth of multiple physical interfaces, or to create a fault-tolerant interface.

2.5.1 Bonding for bandwidth aggregation

Aggregation links utilize multiple interfaces to load-balance traffic to and from the server. There are numerous methods to configure the network interfaces to support load-balancing (sometimes referred to as active/active bonding, because multiple interfaces are actively handling Ethernet traffic).

Round-robin is a method in which traffic is alternated between the active NICs in the team. Gigabit EtherChannel (GEC) is based on Cisco's EtherChannel protocol. Link Aggregation Control Protocol (LACP) is based on the IEEE 802.3ad standard. GEC and LACP require that all interfaces must terminate on the same device. Additional configuration is required on the Ethernet switch to support these protocols.

2.5.2 Fault-tolerant bonding

Redundant bonds are built to protect the traffic flow to and from the server in the event of a NIC, switch, or path failure. This type of configuration is sometimes referred to as active/standby. A minimum of one NIC in the server is designated the *active* NIC, and one NIC is designated as *standby.* An IP address is assigned to the logical interface. The MAC address of the primary NIC is usually assigned to the logical interface. The active NIC handles all transmit and receive traffic. In the event that the active NIC is deactivated, the standby NIC will become the active NIC, and continue to process traffic for the server. The MAC address of the logical network interface will be assigned to the newly-activated NIC. The NIC will send out a gratuitous ARP to inform upstream devices of the change in path.

In the IBM BladeCenter environment, fault-tolerant bonds are the preferred bonding method.

NIC Failover

There are numerous ways in which to trigger the changeover of the active NIC and the standby interfaces:

- Link loss detection at the PHY layer

  If the NIC loses the electrical signals that indicate that it is connected to another device, the Media Independent Interface (MII) signal will go down. This is the same concept as unplugging the Ethernet cable from the NIC on the back of a stand-alone server, or from the directly connected switch. This type of failure is easy for the NIC to detect, because the NIC senses the signal. But, what if something happens in the network where the NIC cannot physically sense the change?

- Loss of protocol connectivity

  The loss of protocol connectivity means that the NIC no longer receives frames or packets from other devices in the network. NIC bonding can be configured to trigger a failover in the case of protocol loss, even though the physical connection is still active. Layer 2 sensing is accomplished with ARP; Layer 3 sensing is accomplished with ping. Broadcom LiveLink (R) accomplishes this for Windows servers running the BACS IM.

Fail-over recovery (failback)

When the failure event has recovered, and the network returns to “normal” operations, the bond will stay with the secondary NIC active. This may or may not be the preferred configuration for the servers. There are features that can automatically return the NICs to the configured position.

There are situations in which failback can cause new problems for the server. Sometimes, during startup and Power on Self test (POST) the switch module may turn on and off the blade-facing switch ports as part of its test. This can cause the bonding driver to return to the active NIC, only to have the link drop again, as the switch continues through POST. It is also possible that after POST, all of the switch links may actually be in the up condition, but the Spanning Tree protocol has not yet finished calculating, and the upstream ports are not in FWD mode. This will cause the server to send traffic out the active NIC, only to be dropped at the switch because traffic is not yet allowed upstream.

This situation can cause instability in the network and for the servers. There are ways to handle this situation, as discussed in later sections.

2.5.3 Bonding in the IBM BladeCenter

Due to the physical layout of the BladeCenter not all teaming methods are supported between the blades and the I/O modules. Although IEEE 802.3ad is supported in the BASP, NetXen, and Linux bonding intermediate drivers, this configuration might not be supported on the links between the blades and the Nortel 10 Gb Ethernet Switch Module, because only one 10 Gb dual-port NIC can be installed on a server, even if the server takes more than one slot in the BladeCenter chassis (such as an LS41). IEEE802.3ad is only supported when both links are connected to the same physical switch.

2.5.4 Linux bonding

NIC bonding in Linux is accomplished using the kernel bonding (bonding.o) IM driver. When bonding Ethernet ports in Linux there are different bonding methods (modes) supported by the bonding driver. The most popular bonding modes are:

- *mode 0* Sets a round-robin policy for fault tolerance and load balancing. Transmissions are received and sent out sequentially on each bonded slave interface beginning with the first one available.
- *mode 1* Sets an active-backup policy for fault tolerance. Transmissions are received and sent out on the first available bonded slave interface. Another bonded slave interface is only used if the active bonded slave interface fails.

Mode 1 (active-backup) is the preferred method in a BladeCenter. Using Mode 0 on a large number of blades can cause problems with the upstream switch or router (whether it is a module in a chassis or not) because each blade will send out multiple *gratuitous ARP* messages per second, associating the bond’s IP address with the alternating different MAC address. This can put a significant load on the upstream router. Additionally, the desired bandwidth increase might not be realized.

Note: Replace the sample BROADCAST, IPADDR, NETMASK and NETWORK values with the appropriate values for your network.

NIC teaming in Windows using the NetXen intermediate driver

The NetXen Teaming driver supports two basic teaming modes; fault tolerant (active/standby) and load balancing with fault tolerance (active/active).

The NetXen adapter provides three fault tolerant options:

- Manual: This setting allows change from a primary NIC to a secondary NIC (only when Switch Now is clicked).
- Fail on Fault: This setting automatically switches from a primary NIC to a secondary NIC when the primary NIC fails.
- Smart Switch: This setting lets a member of a team be selected as the preferred Primary Smart Switch NIC. As long as this NIC is operational, it is always the active NIC. If the NIC fails and it is eventually restored or replaced, it automatically resumes its status as the active NIC. Smart Switch is the recommended choice for fault tolerance. The driver also allows for the configuration of the primary NIC.

Load balancing controls in the NetXen IM driver allow for the selection of four separate controls:

- Transmit load balancing
- Switch-assisted load balancing
- Balance with MAC address
- Balance with IP address

See 5.5.6, “Configuring the NetXen IM driver” on page 71 for more details.

NIC teaming in Windows using the Broadcom Advanced Control Suite

Broadcom offers three methods of teaming.

Smart Load Balancing (SLB)

SLB offers the ability to utilize multiple physical NICs for load balancing and fault tolerance. No additional configuration is required on the network switch to enable SLB. SLB can be configured on a server with NICs from a non-Broadcom vendor, like the NetXen 10 Gb Ethernet adapter.

SLB also offers the option to assign one NIC as a standby NIC. The standby adapter will process traffic if the primary NIC goes down. BASP only allows for one NIC to be assigned as the standby NIC, no matter how many NICs are members of the Active team.

When configuring SLB, there are two separate options for SLB, Smart Load Balancing and SLB with Auto-fallback.

IEEE 802.3ad link aggregation

This mode allows for the configuration of static and dynamic teaming configuration using the Link Aggregation Control Protocol, as defined in the IEEE 802.3ad standard. This teaming method does not work in a typical BladeCenter, as the two NICs are physically connected to two different switches. This method can be implemented on a blade that has two NICs connected to the same switch, as is the case in the LS41, or an HS21 or LS21 with an add-on blade like the Storage and I/O (SIO) add-on or the memory and I/O (MIO) adapter. In such a case, the internal ports of the Ethernet switch must also be configured for IEEE 802.3ad Link aggregation.

Generic link aggregation (trunking)

This method creates a static team (no protocol negotiation). Generic link aggregation is compatible with Cisco EtherChannel (tm) and Lucent Open Trunk (tm) protocols. In this mode,

the team uses a single MAC address. This method has the same restriction in a BladeCenter as 802.3ad link aggregation.

VLAN support

Broadcom Advanced Control Suite also allows for the creation of VLANs on the teamed interface. A VLAN interface is created like a sub-interface. When creating teamed interfaces, an option to created a VLAN is also given. Packets sent and received on a VLAN interface will be tagged with the IEEE 802.1Q tag included, therefore, the switch interface must also support tagging. There can be only one teamed interface that does not have tagging enabled. This interface will use the VLAN assigned to the port in the switch as the native VLAN or the Port VLAN Identifier (PVID).

VLANs are not supported on non Broadcom NICs, or in teams that have mixed vendor NICs.

LiveLink (tm)

LiveLink is a feature offered by Broadcom to monitor the network path the team uses for transmission. Essentially, Livelink sends pings to a user-defined address. If responses are received, the link is considered up and functioning. If the pinged address does not respond, the path is considered down and will trigger a failover (if defined) to the standby NIC.

See 5.5.9, “Installing the Broadcom BASP teaming driver” on page 78 for more details

2.5.5 Trunk failover for high availability

Trunk failover is a feature of the Ethernet switch in the BladeCenter chassis. It works by shutting down ports directly connected to the configured blade servers when a designated set of upstream links go down. The internal ports are put into *disabled* state, and the servers react as though the cable to the network card on a free-standing server had been unplugged. When the configured external trunks recover, the internal ports are re-enabled.

Trunk failover is intended to prevent the following failure mode, when used as part of a High Availability design (Figure 2-8 on page 17):

- The critical connections between a Nortel 10 Gb Switch Module and upstream switch(es) fail, due to a cable problem or the failure of the upstream switch(es).
- The Nortel 10 Gb Switch Module continues to function, and the server blades continue to send traffic to it.
- The Nortel 10 Gb Switch Module, having lost its upstream connections, has no place to forward the server blades’ traffic and therefore discards it.

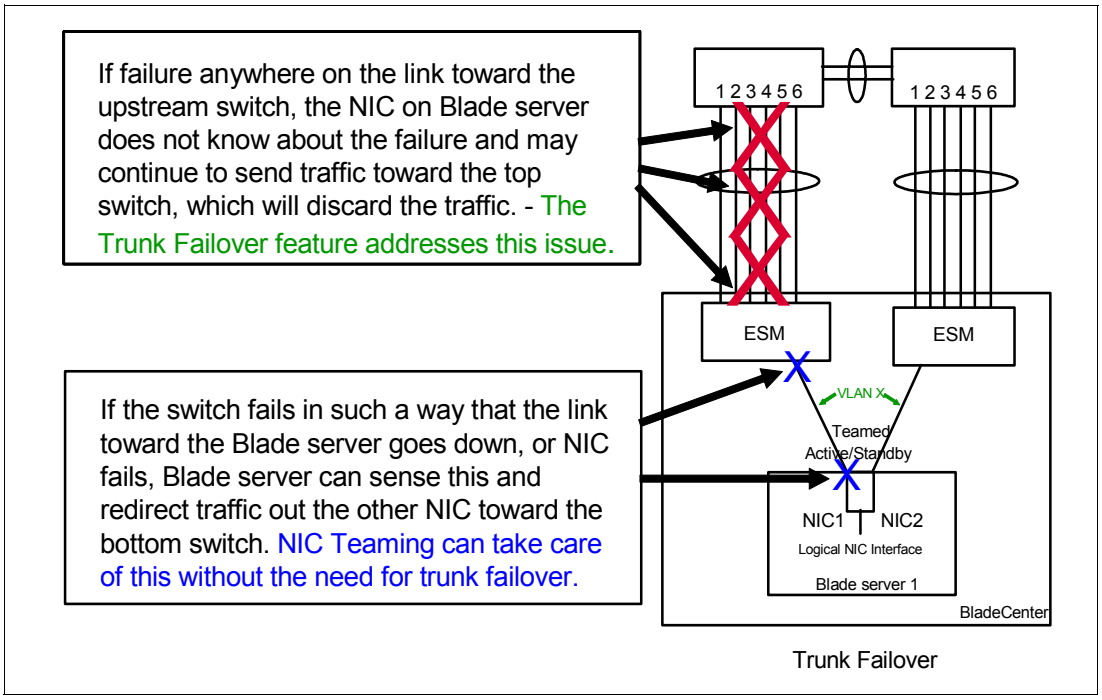


Figure 2-8 Trunk failover protection

Trunk failover is configured on the Nortel 10 Gb Switch Module according to Example 2-1.

Example 2-1 Enabling trunk failover

```
/cfg/l2/failover/on
   trigger <n>
   ena
   amon
      addtrnk <static trunk group number - once for each trunk to be monitored>
      -OR-
      addkey <LACP adminkey value - once for each LACP key to be monitored>
```

If there are multiple trunk groups which are critical upstream connections, such as to multiple upstream switches, then all trunk groups should be added to a trigger condition. Failover will not occur until all of them fail at the same time. It is also possible to create multiple triggers and cause trunk failover for only part of a chassis where multiple VLANs are in use. See the TechTip at http://www.redbooks.ibm.com/abstracts/tips0597?Open for details on how this is done.

In most cases, you should configure trunk failover on all Nortel 10 Gb Switch Modules in the BladeCenter if the server blades are running NIC Teaming.

Refer to Chapter 3, “Nortel 10 Gb Ethernet Switch Module for IBM BladeCenter” on page 23 and Chapter 4, “Nortel Layer 2/3 10 Gigabit Uplink Ethernet Switch Module” on page 51 for more information about configuring trunk failover on Nortel Ethernet switch modules.

2.5.6 Maximizing bonded links in a BladeCenter chassis

When utilizing active/standby NIC bonding, only half of the Ethernet links on the blade server are active at any given time. This implies that only half of the switch modules are being utilized, and half of the available uplink bandwidth is being utilized. A good way to maximize

the bandwidth utilization of the BladeCenter chassis, is to split the active/standby configuration, setting half of the blades to use NIC0 as the active NIC, and the other half of the blades to use NIC1 as the active NIC. Figure 2-9 shows an example of how this configuration might be implemented. Another valid method is to alternate which blades are utilizing the top and bottom switches. This is sometimes referred to as “salt and pepper”.

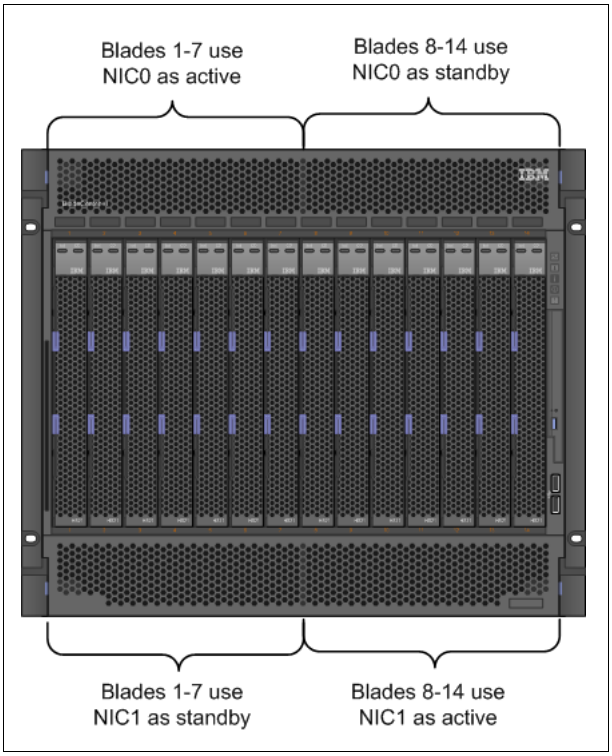


Figure 2-9 Maximizing blade server teamed links in the IBM BladeCenter chassis

2.6 Blade server NIC identification

Every IBM blade server has two Broadcom Ethernet interfaces on the motherboard (technically, LAN on Motherboard, or LOM). These two interfaces will normally enumerate as eth0 and eth1 on Linux, and “Local Area Connection” and “Local Area Connection #2” in Microsoft® Windows.

Sometimes, after various operations with the interfaces within the operating system, it is possible that the adapter names might no longer correspond to the physical interface (for example, it is possible to have Local Area Connection #3 on a blade that has only 2 network interfaces). This can cause confusion while troubleshooting or trying to configure the Ethernet interfaces on the server.

Even though it is not possible to physically trace the connections from the NIC to the switch, there are multiple ways to verify which NIC is connected to each switch module. The method to ensure which NIC is connected where, uses the MAC address of the NIC to identify the connectivity. Some of those methods are detailed in the following sections.

It was once a general rule that when looking at the MAC addresses assigned to the NIC, the even-numbered MAC address was eth0, connected to switch bay one, and the odd-numbered MAC address was eth1, connected to switch bay two. This rule no longer holds true for the

Broadcom adapters on the HS21 or LS21 and LS41 servers. This rule does appear to remain intact for the NetXen 10 Gb Ethernet adapter.

2.6.1 Determine the MAC address from the Management Module

From the Management Module Web interface, select **System Status** → **Hardware VPD**. Newer versions of the AMM firmware will display the MAC addresses of every Ethernet interface on each server, and also those of any adapters or extenders. See Figure 2-10 for an example of the MAC address display.

Bay(s)	Name	MAC Address 1	MAC Address 2
1	HS21A_Std	00:14:5E:EC:42:F8	00:14:5E:EC:42:FA
	HS Exp Card	00:0E:1E:00:13:58	00:0E:1E:00:13:59
5-6	LS41-Ent	00:14:5E:6D:62:64	00:14:5E:B3:5F:16
	Side Board	00:14:5E:6E:16:CA	00:14:5E:6E:16:CC
	HS Exp Card	00:0E:1E:00:13:6A	00:0E:1E:00:13:6B
7	HS21-RH5	00:14:5E:EC:1E:8C	00:14:5E:EC:1E:8E
	HS Exp Card	00:0E:1E:00:13:C0	00:0E:1E:00:13:C1
13	HS21-SLES10	00:14:5E:EC:68:B4	00:14:5E:EC:68:B6
	HS Exp Card	00:0E:1E:00:12:2A	00:0E:1E:00:12:2B
14	HS21-RH4	00:14:5E:EC:68:1E	00:14:5E:EC:68:20
	HS Exp Card	00:0E:1E:00:12:00	00:0E:1E:00:12:01

Figure 2-10 BladeCenter blade MAC addresses in the hardware VPD screen

2.6.2 Determining the MAC address from within the operating system

This section discusses how you determine the MAC address from within the operating system.

Windows MAC address identification

For Windows, navigate to **Start** → **Control Panel** → **Network Connections** → **Local Area Connection**. Right-click the interface to bring up the Status option. From the Support tab of the dialog box, click the **details** button. This will show the physical address of the interface. See Figure 2-11 on page 20 for an example of the dialog box.

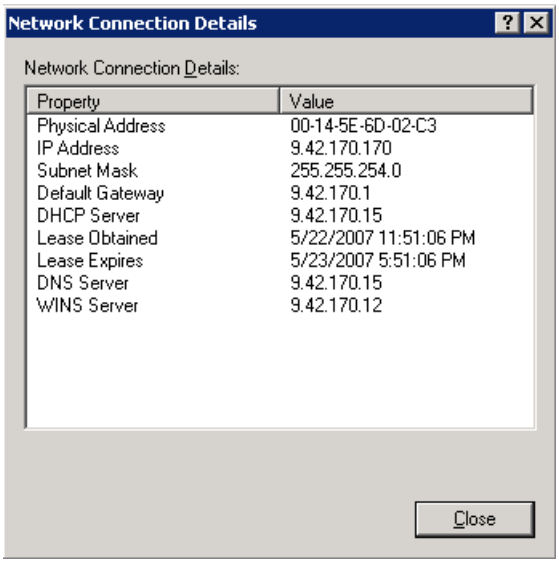


Figure 2-11 Windows dialog box showing the physical address of the network interface

Linux MAC address identification

On a Linux installation, the ifconfig utility will list the HWaddr of each interface (see Example 2-2). it is also possible to gain the information using YAST in SLES or NEAT in Red Hat in the graphical environment.

Example 2-2 Learning the MAC address of the NIC in the Linux OS

```
[root@hs21-rhel4 ~]# ifconfig eth0
eth0      Link encap:Ethernet  HWaddr 00:14:5E:EC:68:1E
          inet addr:9.42.171.100  Bcast:9.42.171.255  Mask:255.255.254.0
          inet6 addr: fe80::214:5eff:feec:681e/64 Scope:Link
          UP BROADCAST RUNNING MULTICAST  MTU:1500  Metric:1
          RX packets:7924546 errors:0 dropped:0 overruns:0 frame:0
          TX packets:9757 errors:0 dropped:0 overruns:0 carrier:0
          collisions:0 txqueuelen:1000
          RX bytes:1145707699 (1.0 GiB)  TX bytes:1051862 (1.0 MiB)
          Interrupt:209 Memory:da000000-da011100
```

Once you know the MAC address of the NIC, it is possible to correctly identify the interface in the switch module.

2.6.3 Determine the MAC address from the switch module

This section discusses how to determine which NIC is connected to which Ethernet switch.

From the Ethernet switch it is possible to determine the MAC address of the interface by checking the forwarding database or MAC address table in the Ethernet switch. The MAC address of the NIC will be listed in the port associated with the server. The forwarding database can be dumped by issuing the **/i/l2/fdb/dump** command on any of the Nortel switch modules.

2.6.4 Determine which NIC is connected to each switch port

From the server's operating system, disable the interface in question. Then check the switch ports in the switch module to identify which interface has changed to the down state.

From the switch module management interface, disable the suspected switch port, and then check the server OS to identify which NIC lost connectivity.

For Linux, **`ethtool eth[x] | grep Link`** will display Link detected yes or no. On Windows servers, the network interface will show a status message indicating that the network cable has been unplugged to the interface.

3

Nortel 10 Gb Ethernet Switch Module for IBM BladeCenter

This section presents an overview of the Nortel 10 Gb Ethernet Switch Module. This includes its physical characteristics, internal architecture, XFP transceivers, and a review of how it differs from the Nortel L2/3 (1 Gb) switch module.

3.1 Product overview

The sections that follow include product description and hardware supported for the Nortel 10 Gb Switch Module.

3.1.1 Product description

The new Nortel 10 Gb Switch Module for the IBM BladeCenter H serves as a high speed switching and routing fabric for the BladeCenter server H or HT chassis. The features and functions of the new 10 Gb switch are essentially the same as those of the Nortel Layer 2/3 GbESM, but at ten times the speed. Like the 1 Gb L2/3 Nortel GbESM, these modules provide Layer 2 switching and Layer 3 routing capabilities.

Figure 3-1 Nortel 10 Gb Ethernet Switch Module for IBM BladeCenter

Figure 3-1 shows a picture of the Nortel 10 Gb Switch Module. Two 10 Gb Ethernet switch modules can reside in the high speed I/O module bays of the BladeCenter H or HT chassis; at present the modules should only be installed in high speed bays 7 and 9 (not 8 or 10) to enable them to communicate with the server blades. In the future it may be possible to deploy four 10 Gb switch modules, occupying bays 7 through 10; the reason this cannot be done at present is that current 10 Gb NIC expansion cards for the server blades access bays 7 and 9 only. These modules can be hot-plugged into a chassis without disrupting normal operations.

The Nortel 10 Gb switch modules connect to the server blades through 14 internal 10 Gb interfaces (server ports) across the BCH/BCHT midplane. There are six external fiber interfaces for outside communication; these external interfaces each require an XFP transceiver.

The switch is managed through two internal 100 Mbps ports for communication to the BladeCenter H Advanced Management Module (AMM). An RS232 serial console management interface and a dedicated 10/100/1000 Ethernet management port are also available.

Full Layer 2 switching and Layer 3 routing provide flexible in-chassis traffic management and security. The Nortel 10 Gb Ethernet Switch module provides full Layer 2 switching with capabilities such as advanced Spanning Tree protocols, Link Aggregation Control, Cisco Etherchannel, and 802.1Q VLANs. Also provided are application delivery and performance

features such as granular QOS (Differentiated Service Code Point 802.1p), Internet Group Management Protocol (IGMP) snooping, and multicast.

In particular, the switch modules support up to 16,384 MAC addresses, 4,096 address resolution protocol (ARP) entries, and up to 2,048 dynamic route entries to ensure a high level of support for a number of users. The IEEE 802.1D Spanning Tree Protocol (STP) support can be enabled or disabled on a per-port basis. Multiple instances of STP are supported (that is, 16 STP groups) and more recent enhancements to STP (802.1w, 802.1s) are also supported. Virtual Local Area Network (VLAN) support includes 802.1Q tagged VLANs and support for IEEE 802.3 support on six external ports for up to three static trunk groups. Dynamic trunking using LACP as well as static trunking is supported.

Adding full Layer 3 routing to the integrated switch module adds more power, flexibility, and security capabilities to the BladeCenter H. With the integrated switch module in the BladeCenter H, network traffic can be managed much more efficiently. Broadcast traffic can be contained in the blade server by placing the 14 blade servers on different subnets while allowing communication between each without using the bandwidth of the external ports to send traffic to and from an external Layer 3 device.

Security features provide added protection for switch configuration data, while packet filtering helps secure and segment sensitive traffic or network access. Support for Simple Network Management Protocol (SNMPv3), Secure Shell (SSHv2), and Hypertext Transfer Protocol over Secure Socket Layer (HTTPS) supply protection for sensitive switch configuration data. Multi-level access and defined access policies help secure the switch against unauthorized management access. Support for Remote Authentication and Remote Authentication Dial-in User Service Protocol (RADIUS), Lightweight Directory Access Protocol (LDAP), and Terminal Access Controller Access Control System (TACACS+) gives enterprises the freedom to use current security databases.

Layer 3 filtering (IP and application type) at line rate in the chassis enhances security and simplifies provisioning. The risk of traffic finding a route to a denied destination is reduced if Layer 3 routing is contained in the switch module in the chassis. Without Layer 3 filtering, several external switches might need configuration to filter traffic to limit access between one server blade and another if the traffic flows through upstream devices.

The following routing standards are supported:

- Routing Information Protocol version 1 (RIPv1), and version 2 (RIPv2)
- Border Gateway Protocol version 4 (BGPv4)
- Open Shortest Path First version 2 (OSPFv2)

The 10 Gb switch module provides cutting edge bandwidth and network flexibility with uplink support for six 10 Gigabit Ethernet ports per switch. The Nortel Networks L2/3 GbESM is designed to be able to route, filter, and queue traffic so that no data is lost, dropped, or delayed. Tests performed by the Tolly Group:
http://www.tolly.com/DecDetail.aspx?DocNumber=206168
show that the switch can support concurrent line speed traffic on all of its ports.

The Nortel 10 Gb Ethernet Switch Module is the only BladeCenter high speed switch module that offers six 10 Gb Ethernet uplink ports for maximum throughput, supporting full Layer 2 through Layer 3 wire-speed packet forwarding for all connections.

High availability support is built in at both Layer 2 and Layer 3 in the Nortel 10 Gb Ethernet Switch Module to reduce single points of failure when it comes to enabling reliability and performance of the network. At Layer 2 Link Aggregation Control (802.3), Rapid Spanning Tree, Fast Uplink Convergence™, Port Fast Forwarding, 802.1Q VLANs, Broadcast Storm Control, and Native Link Failover with NIC teaming are supported. At Layer 3, appropriate

configurations of Virtual Router Redundancy Protocol (VRRP) can allow all switches in the VRRP group to concurrently process traffic by using multiple instances of VRRP. Such configurations enable maximum switch performance while also ensuring seamless failover in the unlikely event of a failure. VRRP Hot Standby is also supported to enable effective use of NIC Teaming in Layer 3 network topologies like Trunk Failover facilitates HA designs with NIC Teaming at Layer 2.

3.1.2 Supported hardware

Table 3-1 lists the BladeCenter chassis which support Nortel 10 Gb Ethernet Switch Module (39Y9267).

Table 3-1 Supported BladeCenter chassis

System name	Machine type	Model
BladeCenter H	8852	All
BladeCenter HT AC	8750	All
BladeCenter HT DC	8740	All

Product shipment group

The items that ship with the switch module are:

- Nortel 10 Gb Ethernet Switch Module (39Y9267)
- Serial Console Cable (43X0509)
- Installation publication, including documentation CD
- Safety flyer
- Software License Agreement

Additional needed hardware

Optical transceivers are sold separately and will be required to use the external ports of the Nortel 10 Gb Switch Module. Figure 3-2 shows a picture of the XFP (10 Gigabit Small Form Factor Pluggable) options.

Figure 3-2 XFP (10 Gigabit Small Form Factor Pluggable) options

Table 3-2 details the available XFP options:

Table 3-2 XFP option characteristics

Module	Description	Max distance	Option number
10 GbASE-SR	850 nm over multi-mode fiber	300 m	32R1877
10 GbASE-LR	1310 nm over single mode fiber	10 km	32R1878

Note: The Nortel 10 Gb Switch Module is designed to supply additional power required for XFPs operating at 1550 nm (40 km and 80 km distances). At the time of this redpaper the current thermal certification is up to the 1310 nm XFP.

3.2 Key differences from the L2/3 GbESM

The following are the key functional differences between the new 10 Gb Ethernet Switch Module and the L2/3 Nortel GbESM (1 Gb switch):

- The 10 Gb switch supports six external 10 Gb ports which will all require fiber connectors and 10 Gb Ethernet XFP transceivers which must be ordered separately. The 1 Gb switch comes with either six copper RJ-45 10/100/1000 ports or six 1 Gb fiber ports with SFP transceivers included and pre-installed.
- The 10 Gb switch module is physically larger than the previous switches for the IBM BladeCenter and can only be installed in the wide High Speed I/O Module. At present these slots are only part of the H and HT chassis.
- The 10 Gb switch module uses the 10 Gb data paths across the midplane of the H and HT chassis, and connects to a 10 Gb Ethernet mezzanine card (CFFh) which must be installed on each blade which will communicate with the switch. These cards are available from NetXen (part# 39Y9271). It is possible that other vendors will provide 10 Gb Ethernet cards which comply with the “Compact Form Factor - horizontal” specification in the future.
- The 10 Gb switch includes an Ethernet management port (EXT7) in addition to the serial console port. This port is not part of or connected to the switching fabric and cannot be used to pass production data; it is only used to manage the switch (using Telnet, ssh, browser, or SNMP). Note that this is different from the 10/100/1000 port on the Nortel 10 Gb Uplink switch (part# 32R1783) which can be used to pass data through the switching fabric.
- The serial console port on the 10 Gb switch uses a mini-USB connector and includes a cable with this connector on one end and a DB9 connector on the other end. The L2/3 GbESM (1 Gb switch) is shipped with a cable with a full-size USB cable on one end and a DB9 on the other end. In both cases these ports and cables use RS-232 signaling and USB devices should never be connected to these ports.
- The 10 Gb switch module supports four default IP gateways (default static routes) which apply to all active VLANs. The *per vlan default gateways* which are included in the Nortel L2/3 and L2/7 GbESMs are not implemented in the 10 Gb switch. The function provided by the *per vlan default gateways* can be implemented if needed through the use of a partial Layer 2 topology, a dynamic L3 routing protocol, or through L3 static routes.
- Software release numbers for the Nortel 10 Gb Ethernet Switch Module do not necessarily correspond exactly to the numbers for the Nortel L2/3 GbESM or those for the Nortel Layer 2/3 10 Gigabit Uplink Ethernet Switch Module. Software revisions for each model are numbered independently. The actual firmware files are coded to identify which module they support and cannot be used in any other models. A diagnostic message will be issued if an attempt is made to load firmware which is intended for a different model.

3.3 Key differences from the Nortel 10 Gb Uplink switch module

The following are the key differences between the new Nortel 10 Gb Switch Module and the Nortel 10 Gb Uplink Switch Module

- The 10 Gb Uplink switch module is built around the traditional form factor and can be installed in any model of IBM BladeCenter chassis. It therefore uses the **1 Gb** paths across the BladeCenter chassis midplane to communicate with the server blades.
- There are three 10 Gb external ports on the 10 Gb Uplink switch module, one of which requires an XFP transceiver for use with multi-mode or single-mode fiber; these are the same XFP and fiber options as the 10 Gb Switch Module. The other two 10 Gb external ports on the 10 Gb Uplink module use copper CX4 cables and have a maximum range of 15 meters.
- There is a 10/100/1000 copper 1 Gb Ethernet port on the 10 Gb Uplink module. This port is a fully functional switch port and can be used to pass data to and from the other switch ports and/or for port mirroring (up to its available bandwidth).

3.4 10 Gb Switch Module architecture

The following sections give a high level overview of the Nortel 10 Gb Switch Module's design.

3.4.1 Block diagram

The Nortel 10 Gb Ethernet switch block diagram is shown in Figure 3-3 below. As shown in the diagram, switching is performed by the Broadcom "Bradley" chip. The Management Processor (MP) on this switch, which supports all switch management, is the PPC 8245 chip.

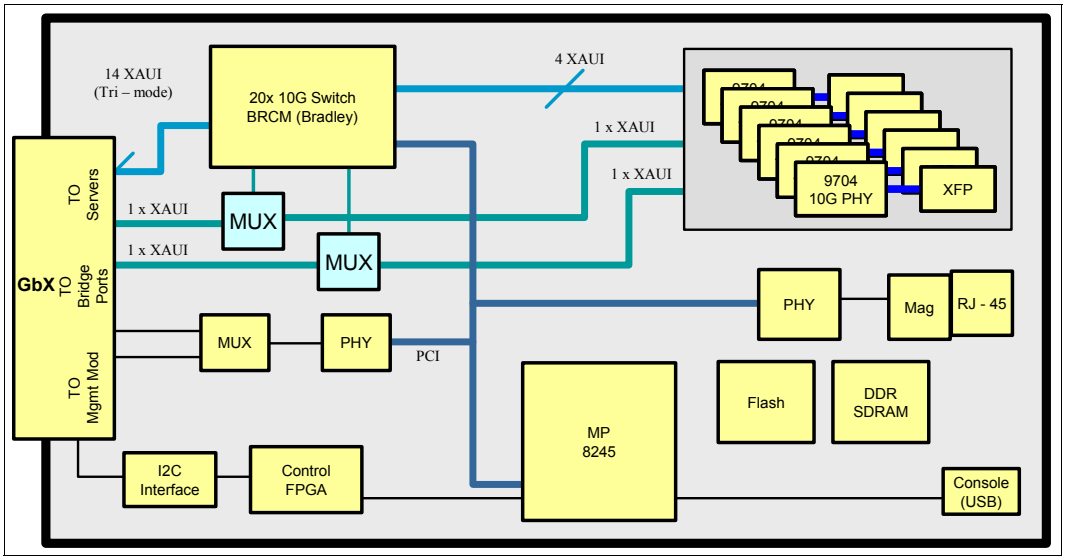


Figure 3-3 Nortel 10 Gb Ethernet Switch Module block diagram

3.4.2 Architecture overview

The following table gives a summary of expected performance level of the Nortel 10 Gb Ethernet Switch Module:

Table 3-3 Nortel 10 Gb Ethernet Switch Module performance expectation

Ports and memory	
Internal ports for blade servers (10 Gbps)	Max of 14
External ports for upstream (10 Gbps)	Max of 6
Internal management ports (100 Mbps)	2
Flash memory	64 MB
Memory dedicated to the management processor	256 MB
Performance	
L2 expected performance	200 Gbps = 10 Gbps * (6 EXT ports + 14 INT ports)
L3 expected performance	200 Gbps

PCI bus

Communication between the MP 8245 and BRCM is handled by a 32-bit PCI bus operating at 66 MHz. The PCI bus is used to configure and control the Broadcom devices. The interface between MP 8245 and BRCM is also used for the following:

- Collecting statistics
- Downloading configuration required for the BRCM and other devices
- Updating configuration information, including filters and VLAN tables

Management interfaces

There are two 1 Gb ports directly connected to the Management Processor, 8245. The Nortel 10 Gb Switch Module firmware always talks to the management module on the same port. Through the use of MM_SELECT_A and MM_SELECT_B signals, the MUX is set to talk to the master management module only.

Note: Only the MGT port associated with the current active management module will show up in /info/link. The MGT port that is associated with the standby management module will show as disabled.

3.5 Switch management

The management paths and rules for connecting to and accessing a BladeCenter Ethernet Switch Module are covered in *Nortel Networks L2/3 Ethernet Switch Module for IBM eServer BladeCenter*, REDP-3586. The following sections review and discuss tools, techniques, and applications that help with the management and deployment of the Nortel 10 Gb Ethernet Switch Module.

3.5.1 Nortel 10 Gb Switch Module management connectivity

In this section, we look at the basic management connectivity and management pathways to the Nortel 10 Gb Ethernet Switch Module, as shown in Figure 3-4 on page 30.

Important: Properly managing the Nortel 10 Gb Ethernet Switch Module in the BladeCenter H actually requires proper management of the Advanced Management Module within the BladeCenter H chassis. In other words, it is virtually impossible to successfully deploy the Nortel 10 Gb Ethernet Switch Module if you do not understand and properly configure certain settings in the Advanced Management Module, as well as the necessary Nortel 10 Gb Ethernet Switch Module configurations.

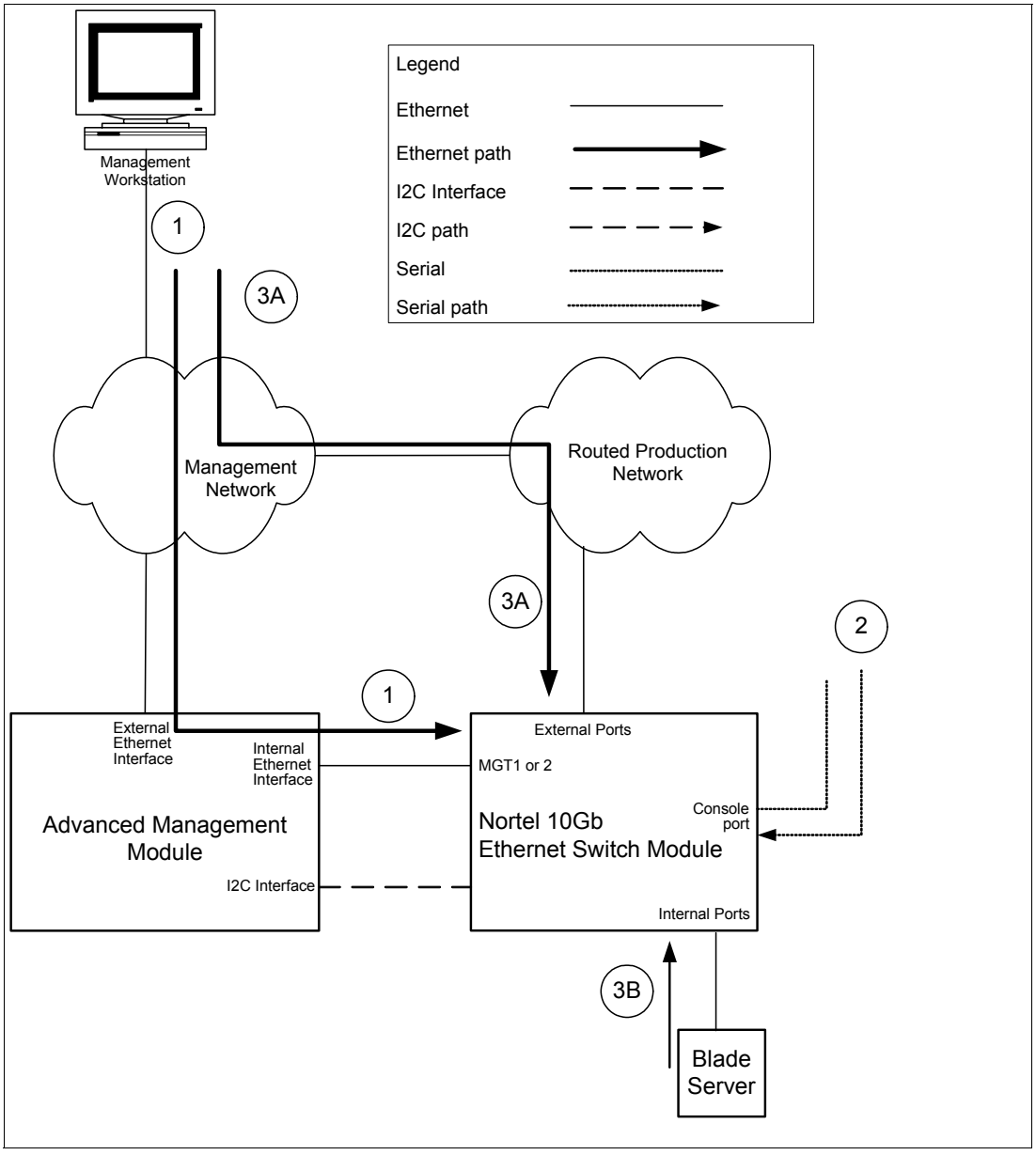


Figure 3-4 Management paths to the Nortel 10 Gb Ethernet Switch Module

Out-of-band management

It is common to provide a (physically) separate management interface for all of the devices and to carry only management traffic. This is referred to as *out-of-band management* and is sometimes a separate Ethernet connection (path 1) or a whole different physical connection such as the console port (path 2).

Advanced Management Module (Path 1)

The BladeCenter H comes with at least one Advanced Management Module. The Advanced Management Module supports an external Ethernet interface, which is used to manage the Blade servers, Ethernet switches, and the Advanced Management Module itself. Within the BladeCenter H, management traffic flows through a different bus, the I2C bus, as shown in the Figure 3-6 on page 34.

On the Nortel 10 Gb Ethernet Switch Module, the Ethernet management (MGT1 and MGT2) ports which connect the switch to the Advanced Management Module are placed in VLAN 4095. It is not possible to change this. It is also not possible to reach VLAN 4095 from any of the other internal or external ports on the switch. This is a deliberate design constraint. It is intended to enforce isolation of the Advanced Management Module network (VLAN) from any other networks (VLANs) that are configured on the switch. This implies that the Blade servers should not be on the same VLAN nor the same IP subnet as the Advanced Management Module. Placing the servers on the same subnet as the Advanced Management Module can have unexpected and undesirable results.

The first step in configuring the Nortel 10 Gb Ethernet Switch Module is to assign the IP address of the MGT ports through the Web interface of the Advanced Management Module (Figure 3-5 on page 32).

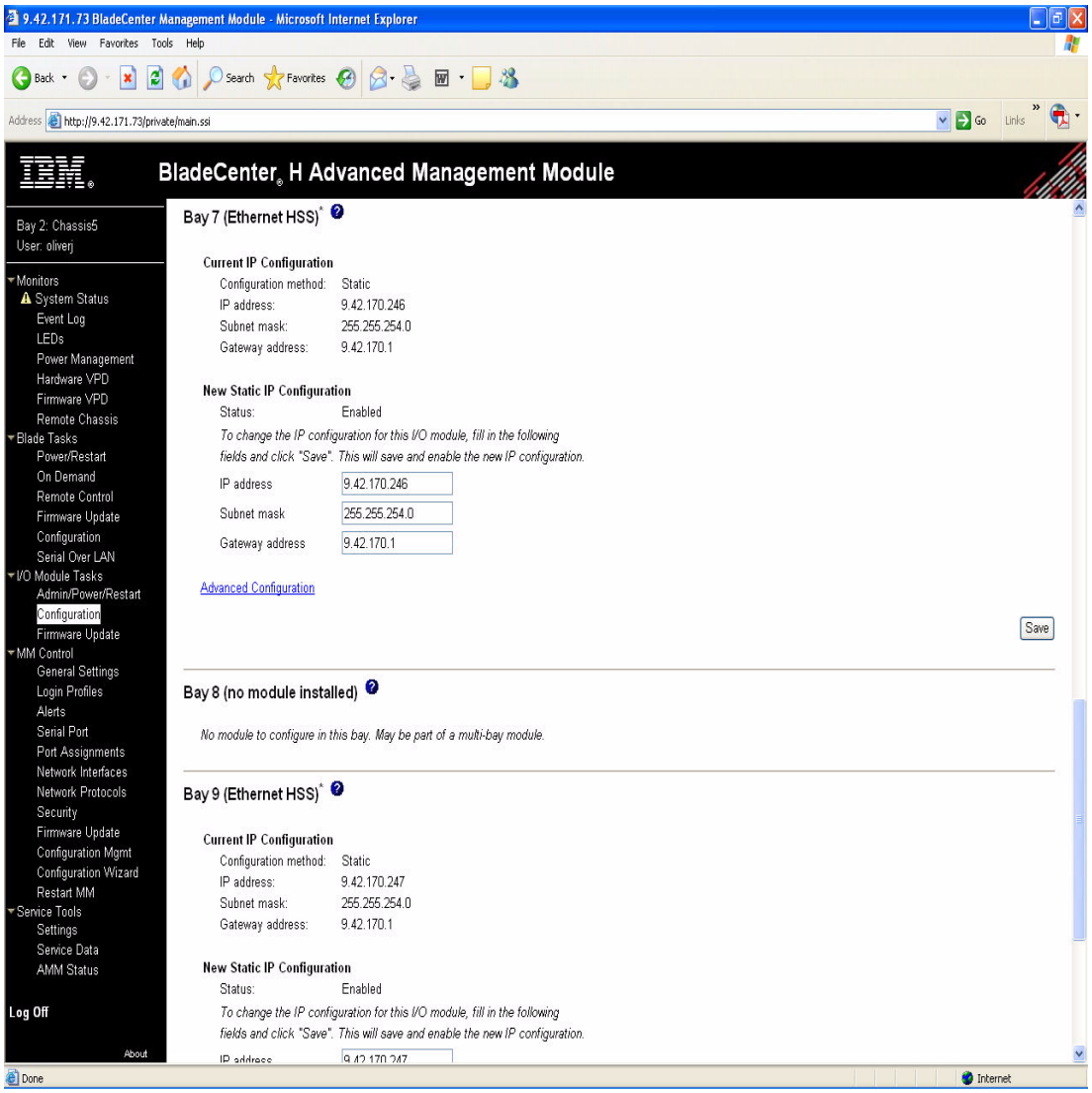


Figure 3-5 Configuring the Nortel MGT port IP address using the Advanced Management Module Web interface

Further configuration of the Nortel 10 Gb Ethernet Switch Module is performed by using Telnet (for the Command Line Interface) or a Web browser (for the Browser Based Interface) and accessing the address of the MGT1 or 2 ports.

Note: It is recommended (and easier) to use a server or mobile computer that is external to the BladeCenter H chassis to perform initial configuration of the Nortel 10 Gb Ethernet Switch Module. The server or mobile computer should be able to open the Web interface of the Advanced Management Module. It then can reach the switch when the switch has an appropriate IP address configured. This address must be within the same subnet as both the internal and external IP addresses of the Advanced Management Module.

Serial port (Path 2)

The serial port is used for out-of-band management of the switch. It is useful to allow access to the CLI when all other paths are not working. It is possible to connect the serial port to a terminal server if desired; this allows out-of-band access to be easily provided to multiple devices.

The console cable that is required to use this port is included with the switch when it is shipped. The cable has a RS232 USB-form plug on one end and a DB-9 plug on the other end. The DB-9 is intended to be attached to a standard serial port such as on a mobile computer or modem. Standard terminal emulation software should be used with these settings: 9600 baud; no parity; 8 data bits; 1 stop bit (9600,N,8,1).

External Management port EXT7 (Path 4)

On the Nortel 10 Gb switch module, there is a 1 Gb Ethernet port which is used exclusively for management of the switch. This is in addition to the MGT ports which communicate through the chassis' Management Module. This port does not carry data to or from the other blade-facing (INTx) or network-facing (EXT1-6) ports. It is not connected to the switching fabric within the 10 GbESM.

Port EXT7 can only be configured to use VLAN 4094, and can only use IP Interface 249 and IP default gateway 253 (the MGT ports use 250 and 254 respectively). EXT7 is different than the MGT port in that the associated IP address and gateway address are ***not*** communicated to the switch from the Management Module, and the IP interface and gateway are *disabled* by default.

In-band management

The second mode of operation that is commonly used is *in-band management*. In this case, the management traffic passes through the data traffic path (the Nortel 10 Gb Ethernet Switch Module External and Internal ports).

External Ethernet ports (Path 3A)

The external ports can be used to provide management access to the switch from outside the BladeCenter H chassis. In order to use this path, the **External management over all ports** item in the Advanced Management Module configuration must be enabled (Figure 3-6 on page 34).

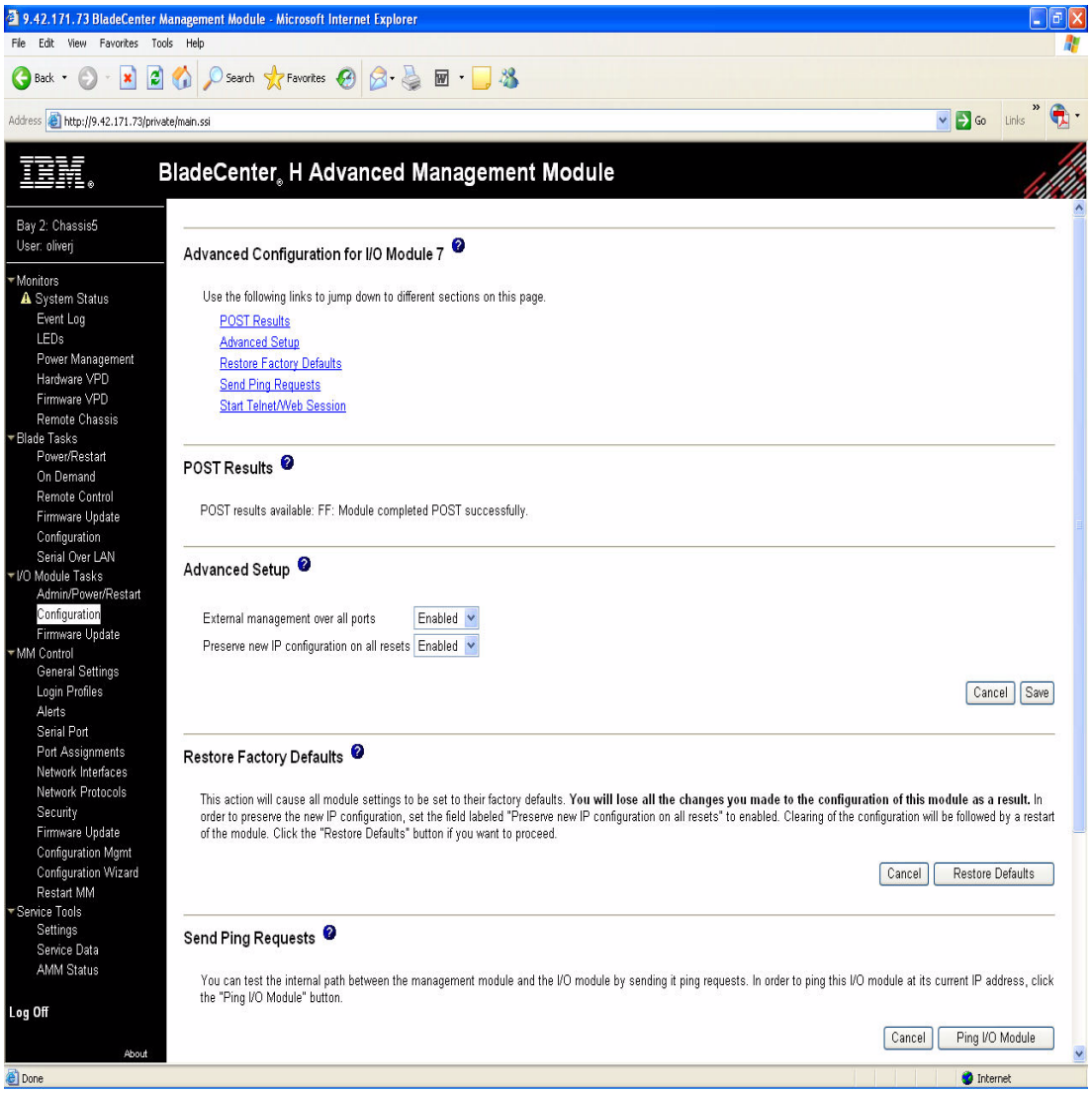


Figure 3-6 Enabling management over all ports using the Advanced Management Module Web interface

Internal Ethernet ports (Path 3B)

The internal ports can be used to provide management access to the switch from the server blades in the same chassis.

In-band management considerations

In order to use in-band management paths, you must configure at least one additional IP address on the Nortel 10 Gb Ethernet Switch Module beyond the address that is provided through the Advanced Management Module and attached to VLAN 4095. This additional IP address should be attached to one of the active VLANs configured on the switch. The following are all supported methods of in-band management that are supported:

- HTTP
- HTTPS
- TELNET
- SSH
- SSL
- SNMP

3.5.2 Nortel 10 Gb Ethernet Switch Module user interface

This section discusses the management interface of the switch module and what each task represents. To configure and manage the switch module, you can use the following interfaces:

- BladeCenter H Advanced Management Module and I2C

  Management functions that are necessary for initial setup are provided through the Advanced Management Module Web interface. I2C is the communication that is used between the Advanced Management Module and Ethernet switch.

- Command-line interface (CLI)

  You can configure and monitor the switch from the CLI, which is accessible through Telnet or SSH from a remote management station. You can also access the CLI through terminal emulation software on a management station directly connected to the switch module console port.

- Browser Based Interface (BBI)

  You can use the Browser Based Interface to manage and monitor the switch using a standard Web browser through HTTP. It provides a graphical means of viewing and configuring the switch's characteristics.

Note: All three of these user interfaces can provide user authentication and secure control through encrypted flows. The standards supported are RADIUS and TACACS+ for the CLI and BBI. LDAP is the standard supported with the Management Module.

BladeCenter H Advanced Management Module and I2C

The Advanced Management Module Web interface is the only mechanism for performing certain management functions, including:

- Configuring the management IP address of the switch
- Enabling or disabling the external ports and management through these ports
- Configuring Power On Self Test (POST) options
- Remotely turning power to the switch on or off

All of these functions use the I2C interface when they need to communicate with the switch module. The use of the Advanced Management Module to configure Ethernet switches is documented in detail in the *Nortel 10 Gb Ethernet Switch Module Installation Guide*.

Command-line interface

The command-line interface CLI is more flexible for configuring the switch than the BBI. It is scriptable, requires less overhead to run, and because it is a Telnet session, it can be run from any operating system (whether or not it is graphical).

Main Menu commands

Figure 3-7 on page 36 shows the Main Menu window. Each of the following commands brings you to a first level submenu:

- The **`stats`** menu gives statistics about the switch.
- The **`cfg`** menu contains all of the configuration options for the switch.
- The **`oper`** menu contains all of the operator commands. Some of these commands can change the state of the switch, but these changes only apply until the next reboot. They are not permanent.

- The **boot** menu contains the commands to control the booting of the switch, from which image to boot, which config to boot, and the **gtimg** and **ptimg** commands for getting and putting firmware files to the switch.
- The **maint** menu contains all of the commands for maintenance of the switch. The commands to manipulate the ARP cache and forwarding database are here, as well as the commands to obtain dumps of the current state of the switch for technical support.

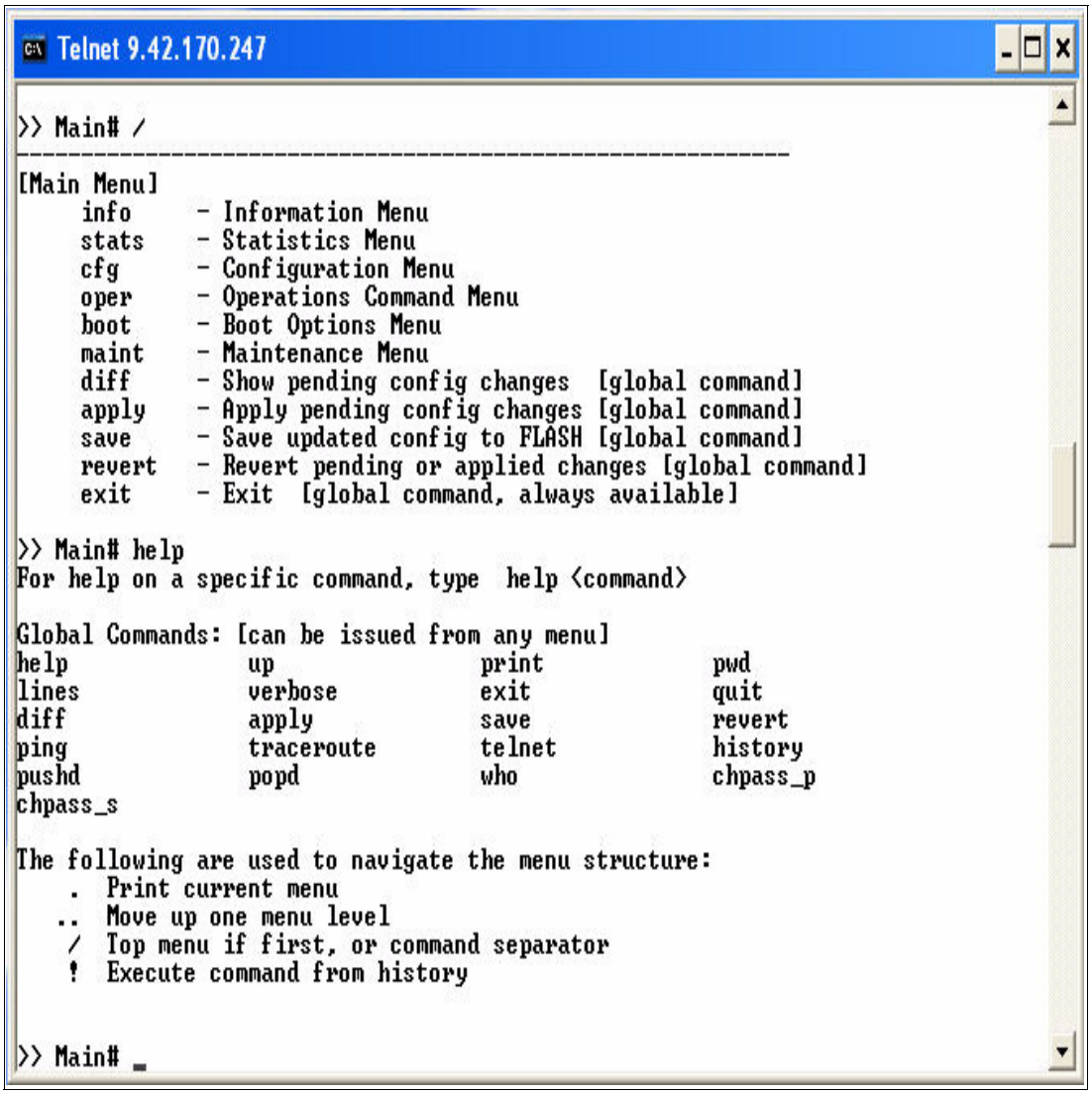


Figure 3-7 CLI menu

Global commands

The remainder of the options on the Main Menu — **diff**, **apply**, **save**, **revert**, and **exit** — are all global commands that work anywhere on the switch. Figure 3-7 also shows what each of the commands does. The **help** command is also global and lists all the global commands, as shown in Figure 3-7.

Navigation commands

There are several commands that are useful in moving from one part of the menu tree to another. The commands are similar to those used in a UNIX® shell:

cd	This command moves you to a given spot in the menu tree. Entering **cd /** always takes you back to the main menu.
pwd	This command displays the current menu path where you are in the menu tree.
.. or **cd ..**	Both of these commands take you up one level in the menu tree.
pushd and **popd**	These commands allow you to manage a stack of menus that you visit frequently.
history	This command displays the last several commands that you entered. You can reuse these commands by typing an exclamation point (**!**) followed by the number of the command as displayed.
quit or **exit**	Both of these commands terminate your session.

Configuration control commands

These commands control the effectiveness of changes to the switch configuration. The general rubric for configuring the switch is EASY:

- E for editing the configuration, typing in your changes
- A for the **apply** command which makes the changes part of the running configuration
- S for the **save** command which writes the changed configuration to flash memory
- Y for yes, which is the answer to the prompt to be sure that you really want to update flash

Additional configuration control commands include the following:

diff	This command displays the differences between the most recent edits and the running configuration.
diff flash	This command displays the differences between the running configuration and its flash copy.
revert	This command discards all changes which have not yet been applied.
revert apply	This command discards applied changes which have not yet been saved to flash.

Additional commands

These are additional commands that facilitate troubleshooting or are otherwise helpful:

ping	Sends ping, Internet Control Message Protocol (ICMP) echo, requests to the specified IP address.
telnet	Opens a Telnet session to the designated IP address.

Note: Both the **telnet** and **ping** commands require an extra switch in the syntax to identify which interface on the switch module the command will run through. The extra switch in the syntax is defined by [-m|-mgt|-e|-ext7|-d|-data]. The following is an example showing both commands and also running a ping to a router connected to the external ports:

```
>> Main# ping ?

ping <hostname>|<IPv4 address> [<tries (1-32)> [<msec-delay>]] [-m|-mgt|-e|-ext7

|-d|-data]


>> Main# telnet ?

telnet <hostname>|<IP address> [port] [-m|-mgt|-e|-ext7|-d|-data]


>> Main# ping 10.10.3.200 -d

Connecting via data port.

[host 10.10.3.200, max tries 5, delay 1000 msec]

10.10.3.200: #1 ok, RTT 0 msec.

10.10.3.200: #2 ok, RTT 0 msec.

10.10.3.200: #3 ok, RTT 0 msec.

10.10.3.200: #4 ok, RTT 0 msec.

10.10.3.200: #5 ok, RTT 0 msec.

Ping finished.


>> Main#
```

traceroute	Traces the IP path to a specified IP address.
who	This command shows who is logged on to the switch and from which address.
verbose	Tailors the level of messages displayed on your session.
lines	Controls the number of lines per screen for display purposes.

Upgrading the firmware

To upgrade the firmware on the Nortel 10 Gb Ethernet Switch Module using the CLI, you must use Trivial File Transfer Protocol (TFTP) or File Transfer Protocol (FTP). Upgrading the switch can also be done using the BBI. It is not possible to use the Advanced Management Module menu item for upgrading firmware at this time.

Important: Before updating the firmware, save any configuration changes to the Nortel 10 Gb Ethernet Switch Module. From the Telnet session, enter **apply,** then press Enter**.** Type **save** and press Enter**.** Answer **y** to the prompt that asks to confirm saving to flash. Answer **y** to the prompt that asks if you want to change the boot to the active config block if it appears.

The firmware for the Nortel 10 Gb Ethernet Switch Module is comprised of two code images. These images are the switch boot code and the switch operating system. These files are labeled Boot and OS respectively. A readme.txt file is packaged with the Nortel 10 Gb Switch Module firmware. This readme file details step-by-step the firmware update process through the CLI.

Capturing the current configuration

There are a few ways to capture the current configuration in the CLI. The first is to use a TFTP server to push the configuration file from the switch to the server. However, in some text editors the resulting file is a single long line of text. (We suggest using WordPad.) Although this method requires a TFTP server running in the network, it does work with any Telnet client. To capture the configuration by pushing a file to a TFTP server:

1. Enter **/cfg/ptcfg** at the command line.
2. Enter the IP address of the TFTP server.
3. Enter the filename to which you want to save the file.

A second way to capture the current configuration does not require a TFTP server. This method, however, requires a terminal emulator that can capture text. Example 3-1 uses a Windows Telnet session to capture the text. The commands on the switch are the same for any software, but the steps to set the software to capture the text might be different. If your terminal emulator does not support this, you have to use the TFTP method. Using a Windows Telnet session and issuing the **/cfg/dump** command, the full switch configuration can be dumped.

Example 3-1 Example configuration file dump

```
>> Main# /cfg/dump
script start "Nortel 10Gb Ethernet Switch Module for IBM BladeCenter" 5  /**** D
O NOT EDIT THIS LINE!
/* Configuration dump taken  7:05:05 Wed May 16, 2007
/* Version 1.0.1.1,  Base MAC address 00:17:ef:c1:6f:00
/c/port INT1
        pvid 2
/c/port INT2
        pvid 2
/c/port INT3
        pvid 2
/c/port INT4
        pvid 2
/c/port INT5
        pvid 2
/c/port INT6
        pvid 2
/c/port INT7
        tag ena
        pvid 2
/c/port INT8
```

```
        pvid 2
/c/port INT9
        pvid 2
/c/port INT10
        pvid 2
/c/port INT11
        pvid 2
/c/port INT12
        pvid 2
/c/port INT13
        pvid 2
/c/port INT14
        pvid 2
/c/port EXT1
        tag ena
        pvid 2
/c/port EXT2
        tag ena
        pvid 2
/c/port EXT4
        tag ena
/c/port EXT6
        tag ena
/c/l2/vlan 1
        def EXT1 EXT2 EXT3 EXT4 EXT5 EXT6
/c/l2/vlan 2
        ena
        name "VLAN 2"
        def INT1 INT2 INT3 INT4 INT5 INT6 INT7 INT8 INT9 INT10 INT11 INT12 INT13
 INT14 EXT1 EXT2 EXT4 EXT6
/c/l2/vlan 3
        ena
        name "VLAN 3"
        def INT7 EXT1 EXT2 EXT4 EXT6
/c/l2/mrst/on
/c/l2/mrst/mode mstp
/c/l2/mrst/rev 1/maxhop 20
/c/l2/mrst/name "TENGIG"
/c/l2/mrst/cist/add 1
/c/l2/mrst/cist/port EXT1/prior 64 /cost 1000
/c/l2/mrst/cist/port EXT2/prior 64 /cost 1000
/c/l2/stg 2/clear
/c/l2/stg 2/add 2
/c/l2/stg 2/port EXT1/prior 64 /cost 1000
/c/l2/stg 2/port EXT2/prior 64 /cost 1000
/c/l2/stg 3/clear
/c/l2/stg 3/add 3
/c/l2/stg 3/port EXT1/cost 1000
/c/l2/stg 3/port EXT2/cost 1000
/c/l2/trunk 1
        dis
        add EXT1
        add EXT2
/c/l2/lacp/port EXT1
        mode active
```

```
        adminkey 10
/c/l2/lacp/port EXT2
        mode active
        adminkey 10
/c/l3/if 1
        ena
        addr 10.10.10.2
        mask 255.255.255.0
        broad 10.10.10.255
        vlan 2
/c/l3/if 2
        ena
        addr 10.10.3.2
        mask 255.255.255.0
        broad 10.10.3.255
        vlan 3
/c/l3/gw 1
        ena
        addr 10.10.10.200
/
script end  /**** DO NOT EDIT THIS LINE!

>>
Configuration#
```

Configuring user accounts

This section describes the user accounts on the switch. There are multiple modes of authentication which are supported on the Nortel 10 Gb Ethernet Switch Module:

- The default mode is to support passwords without individual user IDs. When accessing the CLI in this mode, there is only a single prompt to enter the password. Table 3-4 lists the three passwords that are supported.

Table 3-4 Description of default user accounts

User account	Description and Tasks performed	Default Password
User	Can view switch statistics but cannot make changes.	user
Operator	The Operator account manages all functions of the switch but cannot make permanent changes to the switch configuration.	oper
Administrator	Administrator is the super-user account and has complete access to all menus, information, and configuration commands on the switch.	admin

- Local mode allows the definition of individual user IDs with associated authority levels and passwords. This is configured in the /cfg/sys/access/user menu. For example, Example 3-2 shows the configuration that is necessary to create the BladeCenter H default USERID and PASSW0RD account as an administrator.

Example 3-2 Creation of USERID account as administrator of the switch module

```
>> Main# /c/sys/access/user
------------------------------------------------------------
[User Access Control Menu]
     uid      - User ID Menu
```

```
     eject    - Eject user
     usrpw    - Set user password (user)
     opw      - Set operator password (oper)
     admpw    - Set administrator password (admin)
     strongpw - Strong password menu
     cur      - Display current user status

>> User Access Control# uid
Enter User ID: (1-10) 1
------------------------------------------------------------
[User ID 1  Menu]
     cos      - Set class of service
     name     - Set user name
     pswd     - Set user password
     ena      - Enable user ID
     dis      - Disable user ID
     del      - Delete user ID
     cur      - Display current user configuration

>> User ID 1 # name USERID
Current user name:
New user name:     USERID

>> User ID 1 # cos
Current COS:            user
Enter new COS:          admin

>> User ID 1 # pswd
Changing USERID password; validation required:
Enter current admin password:
Enter new USERID password:
Re-enter new USERID password:
New USERID password accepted.

>> User ID 1 # ena
Current status: disabled
New status:     enabled

>> User ID 1 #
```

- The third authentication mode supported on the Nortel 10 Gb Ethernet Switch Module is using an external authentication server. RADIUS and TACACS+ servers are both supported. These are configured in the /cfg/sys/radius and /cfg/sys/tacacs+ menus respectively. More detail is available in the *Alteon OS 21.0 Application Guide*.

Note: When you attempt to access the switch through the Web interface, you are prompted to enter a user name and password. For all the default users on the switch, the user name and password are the same string by default.

Mode switching to Industry Standard CLI, ISCLI

The first time you start the Nortel 10 Gb Switch Module, it boots into the Alteon OS CLI. To transition to the ISCLI, enter the following command:

```
/boot/mode ISCLI
```

The switch must now be reset to finish the transition to ISCLI mode. The following is an example of an ISCLI configuration file:

Example 3-3 Example configuration file dump using the ISCLI

```
Router>en

Enable privilege granted.
Router#show run
Current configuration:
!
version 1.0.1.1
!

portchannel 1 port EXT1
portchannel 1 port EXT2
no portchannel 1 enable
!
spanning-tree mstp version 1
spanning-tree mstp name "TENGIG"
spanning-tree mode mst
spanning-tree mstp cist-add-vlan 1
!
interface port EXT1
        spanning-tree mstp cist interface-priority 64
        spanning-tree mstp cist path-cost 1000
!
interface port EXT2
        spanning-tree mstp cist interface-priority 64
        spanning-tree mstp cist path-cost 1000
!
spanning-tree stp 2 vlan 2

interface port EXT1
        spanning-tree stp 2 priority 64
        spanning-tree stp 2 path-cost 1000
!
interface port EXT2
        spanning-tree stp 2 priority 64
        spanning-tree stp 2 path-cost 1000
!
spanning-tree stp 3 vlan 3

interface port EXT1
        spanning-tree stp 3 path-cost 1000
!
interface port EXT2
        spanning-tree stp 3 path-cost 1000
!
!
interface port INT1
        pvid 2
!
interface port INT2
        pvid 2
!
interface port INT3
        pvid 2
!
interface port INT4
        pvid 2
!
interface port INT5
        pvid 2
```

```
!
interface port INT6
        pvid 2
!
interface port INT7
        tagging
        pvid 2
!
interface port INT8
        pvid 2
!
interface port INT9
        pvid 2
!
interface port INT10
        pvid 2
!
interface port INT11
        pvid 2
!
interface port INT12
        pvid 2
!
interface port INT13
        pvid 2
!
interface port INT14
        pvid 2
!
interface port EXT1
        tagging
        pvid 2
!
interface port EXT2
        tagging
        pvid 2
!
interface port EXT4
        tagging
!
interface port EXT6
        tagging
!
!
vlan 1
        no member INT1
        no member INT2
        no member INT3
        no member INT4
        no member INT5
        no member INT6
        no member INT7
        no member INT8
        no member INT9
        no member INT10
        no member INT11
        no member INT12
        no member INT13
        no member INT14
        member EXT1
```

```
        member EXT2
        member EXT3
        member EXT4
        member EXT5
        member EXT6
!
vlan 2
        enable
        name "VLAN 2"
        member INT1
        member INT2
        member INT3
        member INT4
        member INT5
        member INT6
        member INT7
        member INT8
        member INT9
        member INT10
        member INT11
        member INT12
        member INT13
        member INT14
        member EXT1
        member EXT2
        member EXT4
        member EXT6
!
vlan 3
        enable
        name "VLAN 3"
        member INT7
        member EXT1
        member EXT2
        member EXT4
        member EXT6
!
interface port EXT1
        lacp mode active
        lacp key 10
!
interface port EXT2
        lacp mode active
        lacp key 10
!
interface ip 1
        ip address 10.10.10.2  255.255.255.0
        enable
        vlan 2
!
interface ip 2
        ip address 10.10.3.2  255.255.255.0
        enable
        vlan 3
!
ip gateway 1 address 10.10.10.200
ip gateway 1 enable
!
end
```

```
Router#
```

To revert back to the Alteon OS CLI, enter the following commands from the ISCLI:

Example 3-4 Commands to restore Alteon command line

```
Router#conf t
Enter configuration commands, one per line.  End with Ctrl/Z.
Router(config)#boot cli-mode aos
Next boot will use mode "aos".
```

Again, the switch must be reset with the **reload** command.

Use the ISCLI Reference Guide for further help using this mode as a means of configuring the Nortel 10 Gb Switch Module.

Browser Based Interface

We now take a brief look at the Browser Based Interface (BBI) on the Nortel 10 Gb Ethernet Switch Module. Almost everything that can be done through the CLI can also be done in the BBI. In the remainder of this paper, more emphasis is placed on configuring the switch using the CLI rather than using the BBI.

The Switch Information panel displays the MAC address of the switch as well as the firmware and hardware versions. Use the following steps to configure the system and contact information:

1. From the Nortel 10 Gb Ethernet Switch Module Web interface, click the folder icon next to Nortel 10 Gb Ethernet Switch Module in the left-hand frame**.**
2. Click the folder icon next to System in the left-hand frame.
3. Click **CONFIGURE** at the top of the page.
4. Click the icon next to General in the list under System. On a window similar to Figure 3-8 on page 47, you see options, such as IP Address and Network Mask fields, that can be configured on this page. Other options on this page include date and time settings, syslog settings (if you have a syslog server), and SNMP settings.

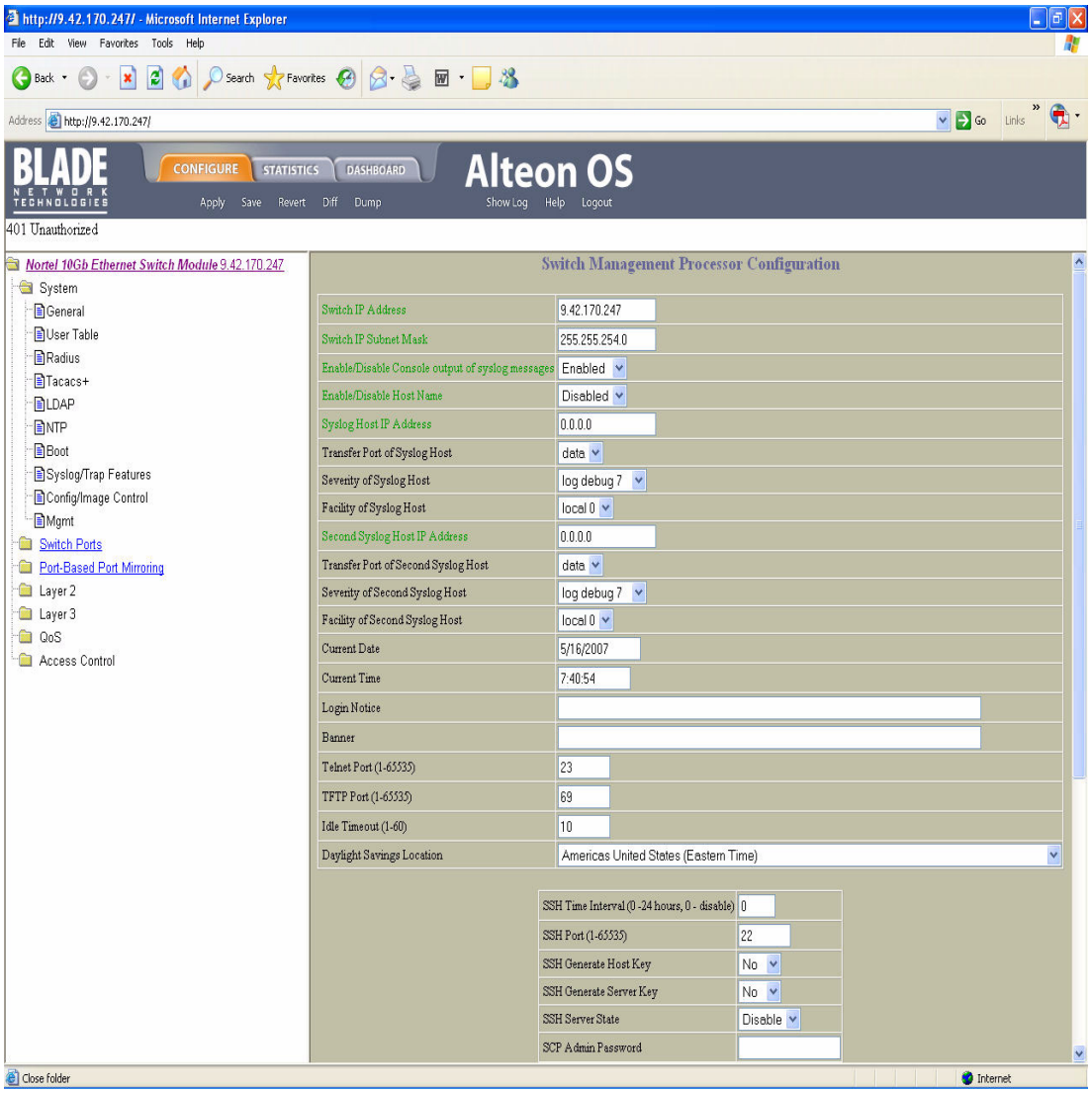


Figure 3-8 Switch information using BBI

You can browse through some of the other links in the left-hand frame to get more familiar with where the configuration options for the switch are located.

5. If you have made any changes to the switch and wish to save them, click **Apply** to apply the changes to the current running config.
6. Click **Save** to save the changes.

> **Note:** It is possible to update the switch firmware through the BBI. This is done from the Config/Image Control link under the System heading.

SNMP management

Two applications providing a graphical way to manage the Nortel 10 Gb Switch Module using SNMP are IBM Director and Blade Harmony.

IBM Director

You can manage and monitor the Nortel 10 Gb Ethernet Switch Module using SNMP through IBM Director. You can also use SNMP-based management systems, such as Tivoli Network Manager. The following SNMP capabilities are supported by the module:

- SNMP management stations can be configured to receive TRAP messages from the switch module. This is configured in the /cfg/sys/ssnmp/ menu. Support is available for SNMPV3 as well as support for SNMP versions 1 and 2.
- SNMP Management Information Base (MIB*)* files are provided with every software image. These files can be imported to the MIB compiler, which is included with IBM Director and other network management products. The MIBs that are provided include Nortel proprietary extensions to the standard MIB1 and MIB2 objects. Both read and write access to these variables can be configured.

Blade Harmony

Blade Harmony is a Java™-based application that runs on the Microsoft Windows operating system. It provides a graphical interface for SNMP-based remote management of Nortel brand switch modules. Some of the benefits of using Blade Harmony include:

- Centralized point of administration of blade switch modules across multiple BladeCenters and racks
- Configuration management and bulk software/image upgrades
- Simplified management of complex configurations such as moving or assigning ports to a large block of filters
- Monitoring and graphing of switch functions for one or many blade switch modules

Note: Blade Harmony supports both the Nortel L2/3 10 Gb Uplink Switch Module and the Nortel 10 Gb Switch Module.

Figure 3-9 on page 49 shows the Blade Harmony user interface managing multiple Nortel switch modules in multiple BladeCenters.

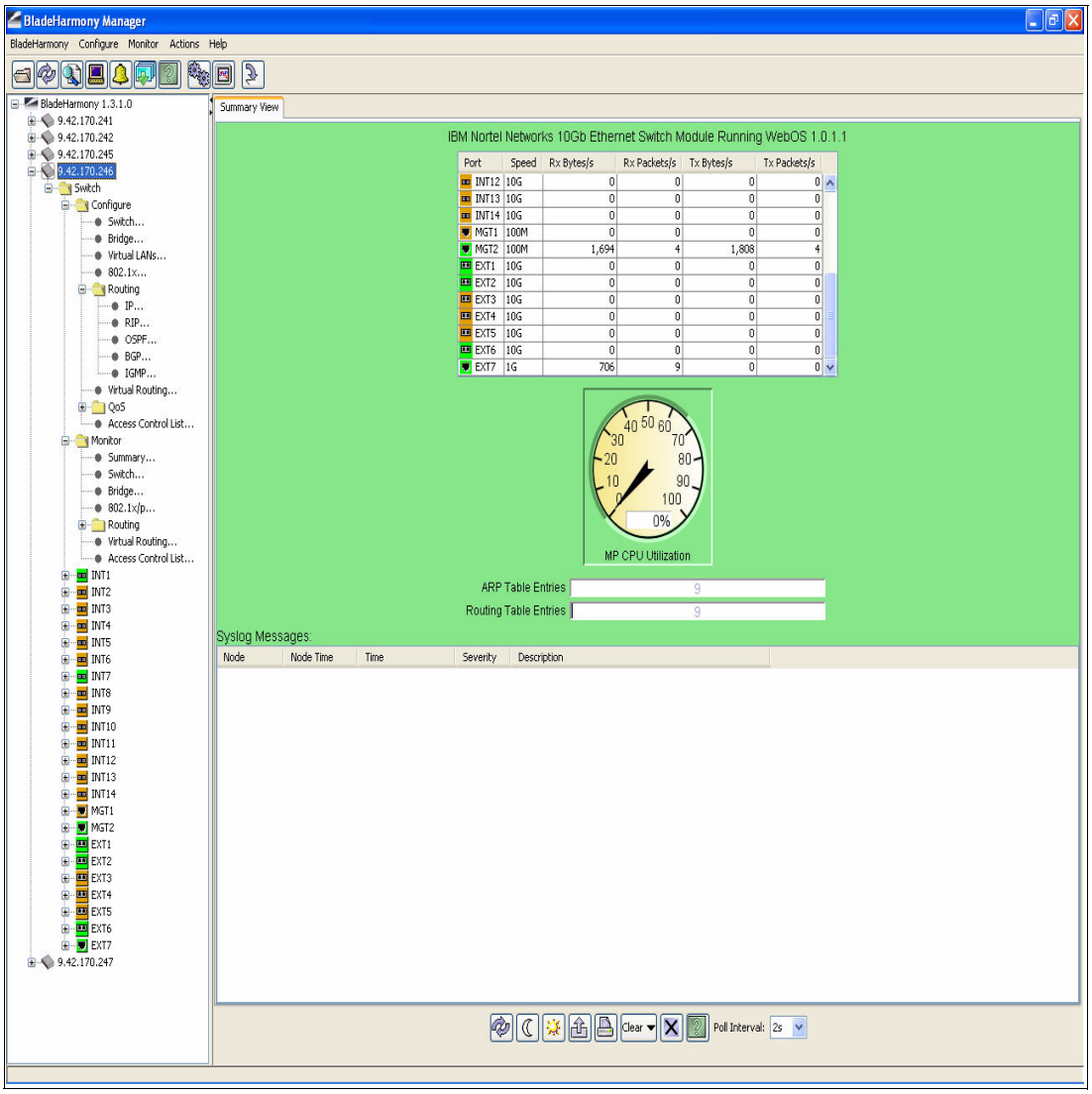


Figure 3-9 Blade Harmony graphical user interface

Note: At the time of this Redpaper, Blade Harmony was available for free download from the Blade Network Technologies Web site. The link to Blade Harmony is:
`http://www.bladenetwork.net/pages/products_blade_harmony.html`

3.5.3 Multiple Nortel 10 Gb Ethernet Switch Modules in a BladeCenter H

If there are two (or more) switches in a single BladeCenter H chassis, the management (MGTx) interfaces of all of the switches are on VLAN 4095. This has the following consequences:

- All of the MGTx IP addresses that are configured through the Advanced Management Module Web interface should be on the same subnet as the Advanced Management Module internal and external port IP addresses (to allow for access through the Advanced Management Module). This configuration also makes it possible to Telnet from one switch module to another across the midplane of the chassis.

- It is not possible to pass substantive data between switch modules across the midplane using the MGTx ports. The Nortel 10 Gb Ethernet Switch Module will not forward data between the MGTx ports and any of the internal (INTx) or external (EXTx) ports. If you want to pass data from one switch module to another, then the modules must be either cabled directly to each other or connected by way of an external switch or router.

The addressing constraints above also apply when the Nortel 10 Gb Switch Module is installed in a BladeCenter chassis along with other switches of any type residing in any of the traditional switch bays or High Speed switch bays.

4

Nortel Layer 2/3 10 Gigabit Uplink Ethernet Switch Module

This section will present an overview of the Nortel Layer 2/3 10 Gigabit Uplink Ethernet Switch Module. This will include its physical characteristics, internal architecture, CX4 copper cabling, XFP transceivers, and a review of how it differs from the Nortel L2/3 (1 Gb) switch module.

4.1 Product overview

The sections that follow include product description and hardware supported for the Nortel 10 Gb Switch Module.

4.1.1 Product description

The Nortel Layer 2/3 10 Gigabit Uplink Ethernet Switch Module for the IBM BladeCenter serves as a high speed switching and routing fabric for the any model of IBM BladeCenter server. This module can use the on-board 1 Gb Ethernet NICs on the blade servers and does not require a daughter board to access switch modules installed in I/O module bays 1 or 2. (Note that Ethernet or iSCSI daughter cards are needed if these switch modules are to be deployed in I/O module bay 3 or 4).

The features and functions of the Nortel Layer 2/3 10 Gigabit Uplink ESM are essentially the same as those of the Nortel Layer 2/3 GbESM, but external (uplink) connectivity is provided at ten times the speed. Like the 1 Gb L2/3 Nortel GbESM, these modules provide Layer 2 switching and Layer 3 routing capabilities.

Figure 4-1 shows a picture of the Nortel Layer 2/3 10 Gigabit Uplink ESM. Up to four 10 Gb Ethernet Uplink switch modules can reside in the I/O module bays of the BladeCenter chassis.

Figure 4-1 Nortel Layer 2/3 10 Gigabit Uplink Ethernet Switch Module

These modules can be hot-plugged into a chassis without disrupting normal operations.

The Nortel Layer 2/3 10 Gigabit Uplink ESM connects to the server blades through 14 internal 1 Gb interfaces (server ports) from each I/O module bay across the BladeCenter midplane. There are three external 10 Gb ports and one 1 Gb external port on the module. Two of the 10 Gb external ports use CX4 copper cables; the remaining one requires an XFP transceiver (SR or LR) for fiber-optic connections. The XFPs supported are the same as those used by the Nortel 10 Gb Ethernet Switch Module. The 1 Gb external port supports connectivity through copper cable with an RJ-45 connector.

The switch is managed through two internal 100 Mbps ports for communication to the BladeCenter Management Module or Advanced Management Module (AMM). An RS232 serial console management interface is also available.

Full Layer 2 switching and Layer 3 routing provide flexible in-chassis traffic management and security. The Nortel Layer 2/3 10 Gigabit Uplink ESM provides full Layer 2 switching with capabilities such as advanced Spanning Tree protocols, Link Aggregation Control, Cisco Etherchannel, and 802.1Q VLANs. Also provided are application delivery and performance features such as granular QOS (Differentiated Service Code Point 802.1p), Internet Group Management Protocol (IGMP) snooping, and multicast.

In particular, the switch modules support up to 16,384 MAC addresses, 4,096 address resolution protocol (ARP) entries, and up to 2,048 dynamic route entries to ensure a high level of support for a number of users. The IEEE 802.1D Spanning Tree Protocol (STP) support can be enabled or disabled on a per-port basis. Multiple instances of STP are supported (that is, 16 STP groups) and more recent enhancements to STP (802.1w, 802.1s) are also supported. Virtual Local Area Network (VLAN) support includes 802.1Q tagged VLANs and support for IEEE 802.3 support on six external ports for up to three static trunk groups. Dynamic trunking using LACP as well as static trunking is supported.

Adding full Layer 3 routing to the integrated switch module adds more power, flexibility, and security capabilities to the IBM BladeCenter. With the integrated switch module in the BladeCenter, network traffic can be managed much more efficiently. Broadcast traffic can be contained in the blade server by placing the 14 blade servers on different subnets while allowing communication between each without using the bandwidth of the external ports to send traffic to and from an external Layer 3 device.

Security features provide added protection for switch configuration data, while packet filtering helps secure and segment sensitive traffic or network access. Support for Simple Network Management Protocol (SNMPv3), Secure Shell (SSHv2), and Hypertext Transfer Protocol over Secure Socket Layer (HTTPS) supply protection for sensitive switch configuration data. Multi-level access and defined access policies help secure the switch against unauthorized management access. Support for Remote Authentication and Remote Authentication Dial-in User Service Protocol (RADIUS), Lightweight Directory Access Protocol (LDAP), and Terminal Access Controller Access Control System (TACACS+) gives enterprises the freedom to use current security databases.

Layer 3 filtering (IP and application type) at line rate in the chassis enhances security and simplifies provisioning. The risk of traffic finding a route to a denied destination is reduced if Layer 3 routing is contained in the switch module in the chassis. Without Layer 3 filtering, several external switches might need configuration to filter traffic to limit access between one server blade and another if the traffic flows through upstream devices.

The following routing standards are supported:

- Routing Information Protocol version 1 (RIPv1), and version 2 (RIPv2)
- Border Gateway Protocol version 4 (BGPv4)
- Open Shortest Path First version 2 (OSPFv2)

The 10 Gb Uplink switch module provides cutting edge bandwidth and network flexibility with uplink support for three 10 Gigabit Ethernet ports per switch. The Nortel Layer 2/3 10 Gigabit Uplink ESMs are designed to be able to route, filter, and queue traffic so that no data is lost, dropped, or delayed. Tests performed by the Tolly Group:
http://www.tolly.com/DocDetail.aspx?DocNumber=206168
show that the switch can support concurrent line speed traffic on all of its ports.

High availability support is built in at both Layer 2 and Layer 3 in the Nortel 10 Gb Ethernet Uplink Switch Module to reduce single points of failure when it comes to enabling reliability and performance of the network.

At Layer 2 Link Aggregation Control (802.3), Rapid Spanning Tree, Fast Uplink Convergence, Port Fast Forwarding, 802.1Q VLANs, Broadcast Storm Control, and Native Link Failover with NIC teaming are supported.

At Layer 3, appropriate configurations of Virtual Router Redundancy Protocol (VRRP) can allow all switches in the VRRP group to concurrently process traffic by using multiple instances of VRRP. Such configurations enable maximum switch performance while also ensuring seamless failover in the unlikely event of a failure. VRRP Hot Standby is also supported to enable effective use of NIC Teaming in Layer 3 network topologies like Trunk Failover facilitates HA designs with NIC Teaming at Layer 2.

4.1.2 Supported hardware

Table 4-1 lists the following BladeCenter chassis which support Nortel Layer 2/3 10 Gigabit Uplink ESM.

Table 4-1 Supported BladeCenter chassis

System name	Machine type	Model
BladeCenter	8677	All
BladeCenter T (AC Power)	8730	All
BladeCenter T (DC Power)	8720	All
BladeCenter H	8852	All
BladeCenter HT (AC Power)	8750	All
BladeCenter HT (DC Power)	8740	All

Product shipment group

The items that ship with the switch module are:

- Nortel 10 Gb Ethernet Uplink Switch Module (32R1783)
- Serial Console Cable (02R9362)
- Installation publication, including documentation CD
- Safety flyer
- Software License Agreement

Additional needed hardware

A single optical transceiver is needed for the 10 Gb fiber uplink port. CX4 cables are needed for the two 10 Gb copper uplink ports.

Optical transceivers

Optical transceivers are sold separately and will be required to use the 10 Gb optical external port of the Nortel 10 Gb Uplink Switch Module. Figure 4-2 on page 55 shows a picture of the XFP (10 Gigabit Small Form Factor Pluggable) options.

Figure 4-2 XFP (10 Gigabit Small Form Factor Pluggable) options

Table 4-2 details the available XFP options:

Table 4-2 XFP option characteristics

Module	Description	Max distance	Option number
10 GbASE-SR	850 nm over multi-mode fiber	300 m	32R1877
10 GbASE-LR	1310 nm over single mode fiber	10 km	32R1878

CX4 cabling

CX4 cables are sold separately and are required to use the 10 Gb copper external ports of the Nortel Layer 2/3 10 Gigabit Uplink ESM. IBM sells 1.5 meter and 3 meter cables, which are shown in Table 4-3.

Table 4-3 CX4 cable option characteristics

Cable type	Description	Option number
CX4	1.5 m	32R1937
CX4	3.0 m	32R1941

Note: Cables up to 15 meters in length are supported and can be purchased from other cable vendors.

Figure 4-3 shows the CX4 cable type.

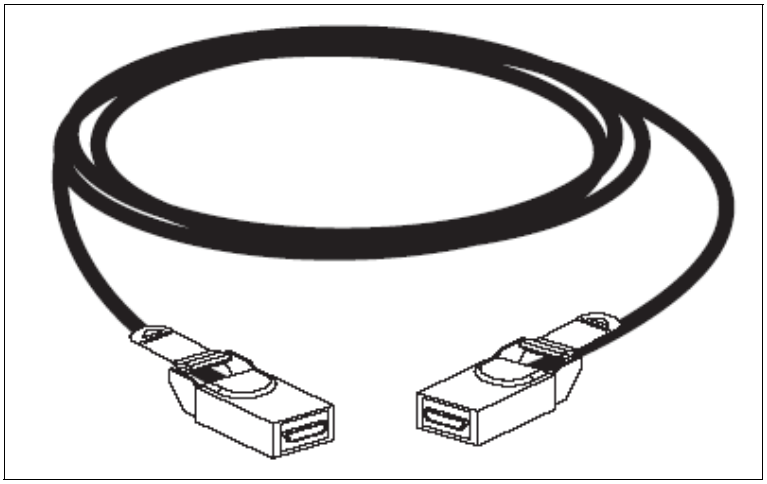


Figure 4-3 CX4 cable

Figure 4-4 shows the 10 Gb CX4 connector.

Figure 4-4 CX4 connector

4.2 Key differences from the L2/3 GbESM

The following are the key functional differences between the new 10 Gb Ethernet Switch Module and the L2/3 Nortel GbESM (1 Gb switch):

- The Nortel Layer 2/3 10 Gigabit Uplink ESM provides three external 10 Gb ports, one of which will require fiber connectors and 10 Gb Ethernet XFP transceivers which must be ordered separately and the remaining two require CX4 cabling as described in section 4.1 above. The module also provides a single copper RJ-45 1 Gb Ethernet port. The Nortel Layer 2/3 1 Gb ESM comes with either six copper RJ-45 10/100/1000 ports or six 1 Gb fiber ports with SFP transceivers included and pre-installed.
- Software release numbers for the Nortel Layer 2/3 10 Gigabit Uplink ESM do not necessarily correspond exactly to the numbers for the Nortel L2/3 GbESM or those for the Nortel 10 Gb Ethernet Switch Module. Software revisions for each model are numbered independently. The actual firmware files are coded to identify which module they support and cannot be used in any other models. A diagnostic message will be issued if an attempt is made to load firmware which is intended for a different model.
- The 10 Gb Uplink switch module supports four default IP gateways (default static routes) which apply to all active VLANs. The *per vlan default gateways* which are included in the Nortel L2/3 and L2/7 GbESMs are not implemented in the 10 Gb switch. The function provided by the *per vlan default gateways* can be implemented if needed through the use of a partial Layer 2 topology, a dynamic L3 routing protocol, or through L3 static routes.

4.3 10 Gb Switch Module architecture

The following sections give a high level overview of the Nortel 10 Gb Switch Module's design.

4.3.1 Block diagram

The Nortel 10 Gb Ethernet switch block diagram is shown in Figure 4-5 on page 57. As shown in the diagram, switching is performed by the Broadcom " the BCM part number or a generic label" chip. The Management Processor (MP) on this switch, which supports all switch management, is the PPC 8245 chip.

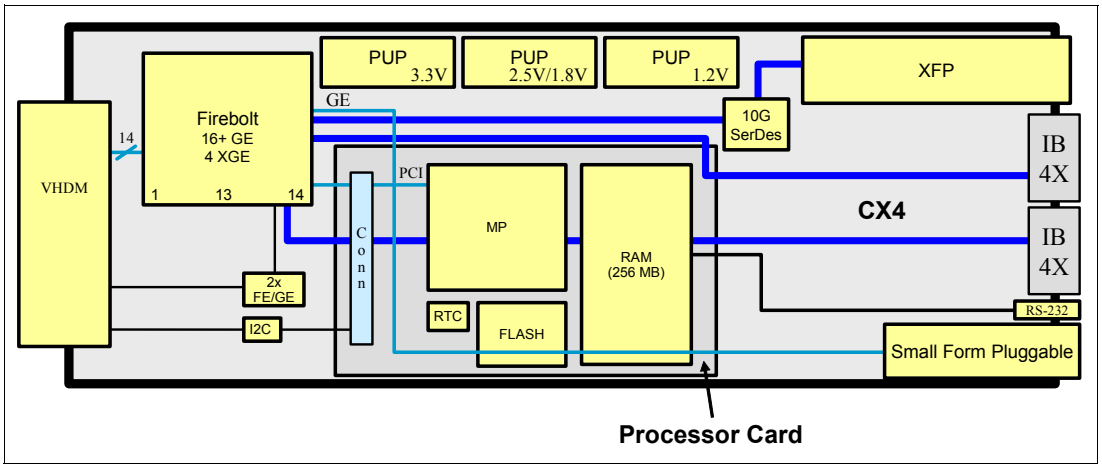


Figure 4-5 Nortel 10 Gb Ethernet Uplink Switch Module block diagram

4.3.2 Architecture overview

The following table gives a summary of expected performance level of the Nortel Layer 2/3 10 Gigabit Uplink ESM:

Table 4-4 Nortel 10 Gb Ethernet Uplink Switch Module performance expectation

Ports and memory	
Internal ports for blade servers (10 Gbps)	Max of 14
External ports for upstream (10 Gbps)	Max of 3 (two CX4 copper and one XFP fiber)
Internal management ports (100 Mbps)	2
Flash memory	64 MB- check
Memory dedicated to the management processor	256 MB
Performance	
L2 expected performance	45 Gbps = 1 Gbps * 15 ports + 10 Gbps * 3 ports
L3 expected performance	45 Gbps

Management interfaces

There are two 1 Gb ports directly connected to the Management Processor, 8245. The Nortel 10 Gb Uplink Switch Module firmware always talks to the management module on the same port. Through the use of MM_SELECT_A and MM_SELECT_B signals, the MUX is set to talk to the master management module only.

Note: Only the MGT port associated with the current active management module will show up in /info/link. The MGT port that is associated with the standby management module will show as disabled. This is the same behavior as the Nortel 10 Gb Switch Module.

4.4 Switch management

The management paths and rules for connecting to and accessing a BladeCenter Ethernet Switch Module are covered in *Nortel Networks L2/3 Ethernet Switch Module for IBM eServer BladeCenter*, REDP-3586. Management paths and rules have also been reviewed in Chapter 3. Techniques and applications aiding management and deployment of the switch are also covered in Chapter 3. These are fundamentally the same for the Nortel Layer 2/3 10 Gb Uplink ESM.

5

NetXen card installation and configuration

This chapter discusses the NetXen 10 Gb Ethernet adapter for IBM BladeCenter.

5.1 Physical card

NetXen 10 Gb Ethernet Expansion Card for IBM BladeCenter (Figure 5-1) is a new high-speed expansion card especially designed to interface with the Nortel 10 Gb Ethernet Switch Module for the IBM BladeCenter H and HT chassis.

The 10 Gb Ethernet Expansion Card provides the following:

- Two 10 Gigabit SERDES interfaces from each blade to the primary and redundant high-speed module switch bay locations (switch bays 7 and 9)
- PCIe-based BladeCenter high-speed expansion card (CFFh) that can be simultaneously paired with CFFv cards.

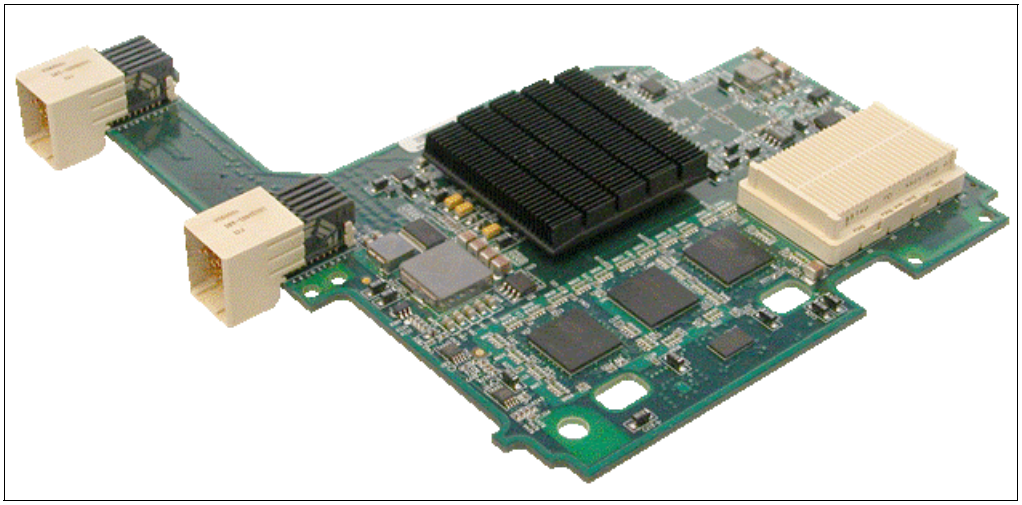


Figure 5-1 A picture of the NetXen 10 Gb Ethernet Expansion Card for IBM BladeCenter

5.2 Architectural overview

This section describes the high level architecture of the NetXen 10 Gb Ethernet Expansion Card (Figure 5-2 on page 61) and its implementation as a Combo Form Factor horizontal (CFFh) expansion card. The major components on the card include the NX2031 ASIC, memory to hold the packet buffers, and Flash to store the configuration and the firmware.

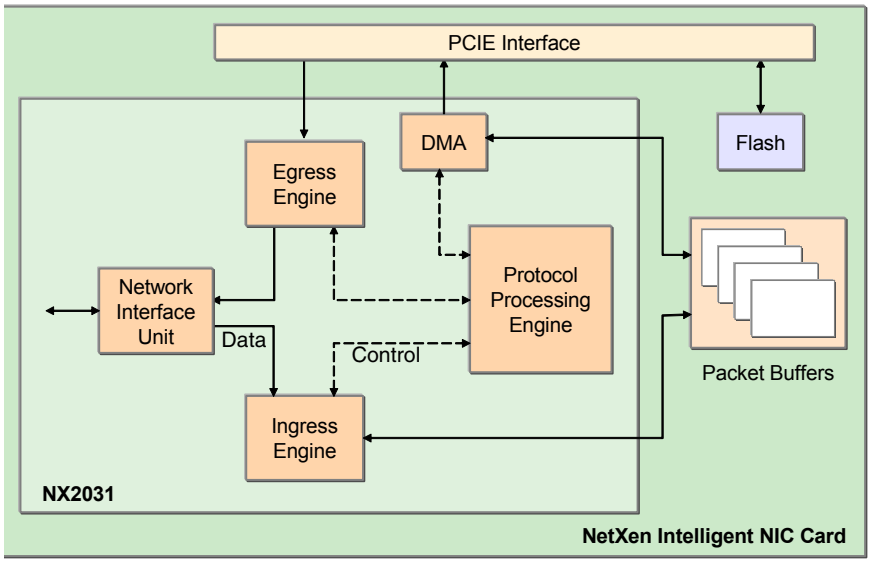


Figure 5-2 NetXen 10 Gb Ethernet Expansion Card architecture

5.2.1 High level architecture of the NetXen 10 Gb card

The Network Interface Unit (NIU) has 2x10G MAC and the SerDes. It is responsible for the PHY interface. The NIU handles incoming and outgoing Unicast and Multicast packets. It has the MAC filters to receive Multicast packets and supports 16 Unicast and 48 Multicast MAC addresses.

Transmit MAC

The transmit MAC performs the following functions:

- Preamble and Start of Frame Delimiter (SFD)
- Padding
- CRC Generation
- Inter packet Gap (IPG)
- MAC control frame generation
- Error sequence

Receive MAC

The receive MAC performs the following functions:

- Preamable and SFD removal
- CRC removal
- Unicast address filter - The Unicast designation address of the receive packet is compared with stored MAC address. The unmatched packets are dropped.
- Multicast address filter
- Broadcast address filter
- Frame Validity Checks
- Pause frame detection
- Frame synchronization and sequence error detection

The receive MAC has link fault signaling state machine to process received sequence error.

Egress Engine (EE)

The EE is the hardware block that transmits the packets from either the host or the onboard memory.

Ingress Engine (IE)

The IE receives the packets and processes them for DMA to the host.

DMA

On-board DMA engines provide efficient data transfers across the PCIe bus. The main features of the DMA controller are:

- 64-bit addressing
- Supports multiple outstanding DMA operations that are executed in parallel
- DMA operations executed in both directions: host to NX2031, and NX2031 to host
- Supports full scatter/gather mode
- DMA operations can be chained together, allowing the NX2031 to perform a large number of contiguous DMA operations

Protocol Processing Engine

The protocol processing engine is a flexible, programmable engine that can handle complex processing operations.

PCI Express Interface

The PCI Express interface is PCIE 1.0 compliant. The PCIE interface supports message based interrupts MSI and MSI-X, as well as traditional INTA interrupts. The interface provides full support for expansion ROM (PCIE boot) and Vital Product Data (VPD).

5.3 Supported platforms

The following two sections list the hardware and software platforms that are compatible with the NetXen 10 Gb Ethernet Expansion Card for IBM eServer BladeCenter.

5.3.1 Servers

The NetXen 10 Gb Ethernet Expansion Card for IBM eServer BladeCenter (highlighted in Figure 5-3 on page 63) is supported on the following IBM BladeCenter blades inserted in a BladeCenter H or BladeCenter HT chassis with the Nortel 10 Gb Ethernet Switch Module.

- BladeCenter HS21 8853-xxx
- BladeCenter HS21 XM 7995-xxx
- BladeCenter LS21 7971-xxx
- BladeCenter LS41 7972-xxx
- IBM BladeCenter Storage and I/O expansion blade (39R7563)
- IBM BladeCenter Memory and I/O blade (42C1600)
- Future blades

Note: The NetXen 10 Gb Ethernet Expansion Card for IBM eServer BladeCenter physically fits in the JS21, but at the time of writing no drivers for AIX® or Linux on PowerPC are available.

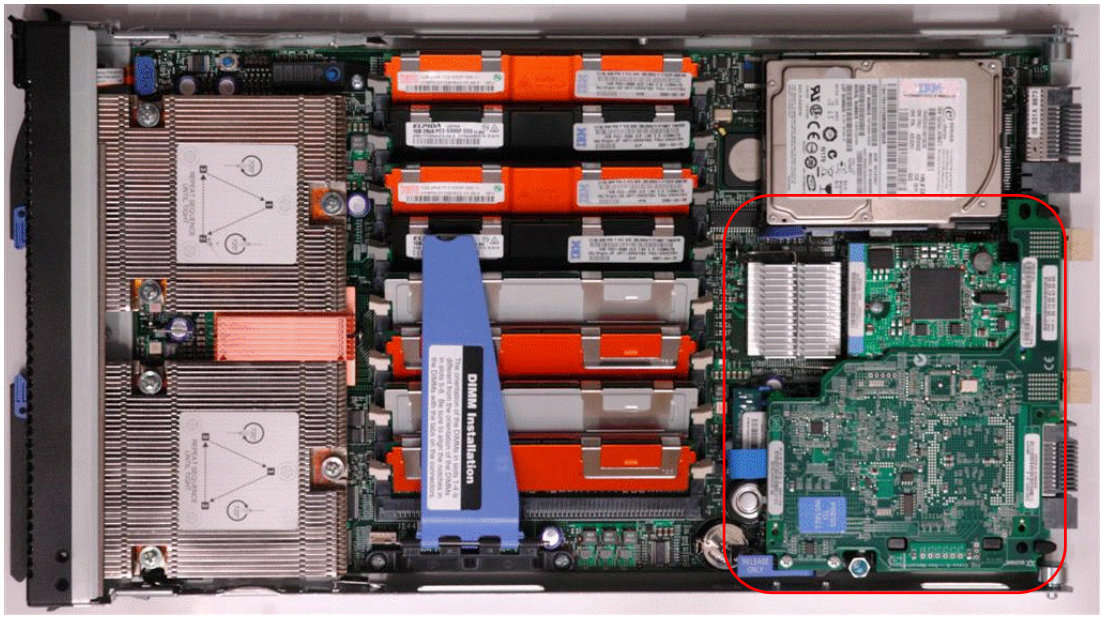


Figure 5-3 Combo Form Factor Horizontal card installed in HS21XM server

5.3.2 Operating system support

The NetXen 10 Gb Ethernet Expansion Card for IBM eServer BladeCenter is supported under the following operating systems:

- Microsoft Windows 2000 (Server and Advanced Server)
- Microsoft Windows 2003/2003 R2 (Standard, Enterprise, Web)
- Microsoft Windows 2003/2003 R2 x64 Edition (Standard, Enterprise)
- Red Hat Enterprise Linux 4 for x86 (AS, ES, WS)
- Red Hat Enterprise Linux 4 for AMD/EM64T (AS, ES, WS)
- Red Hat Enterprise Linux 5 Server
- SUSE LINUX Enterprise Server 9 for x86
- SUSE LINUX Enterprise Server 9 for AMD64/EM64T
- SUSE LINUX Enterprise Server 10 for x86
- SUSE LINUX Enterprise Server 10 for AMD64/EM64T

Note: Support for VMWare is expected with the release of VMware VI 3.1. Support for Linux on Power and AIX on JS21 blades is planned. None of these Operating Systems were supported at the time this document was written.

5.4 NIC teaming or bonding

Refer to section 2.5, “NIC bonding” on page 13 for more information.

5.5 Windows 2003 SP2

The NetXen 10 Gb Ethernet adapter is supported on Windows Server® 2003 Enterprise, Standard and Web editions. Both 32-bit and 64-bit versions are supported. Service Pack 2 (SP2) must be installed on the operating system prior to installation of the NetXen drivers. Always check the latest readme file for installation instructions, system requirements and caveats when installing a new driver.

5.5.1 Drivers

The NetXen driver is supplied by IBM through the Document IDs below that link to the NetXen Web site. You must register prior to downloading the document. Currently available are Windows and Linux versions, with other drivers in development. Drivers for Windows are available in 32- and 64-bit versions.

The driver package contains:

- NIC Windows driver
- IM Windows driver (teaming driver)
- Firmware
- Diagnostics
- Tools
- Documentation in PDF format

The GA level of the software release for Windows and Linux is: Version 3.4.216.

Note: The software release consists of driver and firmware.

5.5.2 Installing the Windows drivers

Download and unzip the latest driver from the Document ID provided in 5.5.1, “Drivers” on page 64 in a directory on the blade or, if you need to install on multiple blades, create an installation CD.

If this is the first boot after installation of the NetXen 10 Gb Ethernet Expansion Card for IBM eServer BladeCenter, Windows will detect the new hardware on boot. Point the installation dialogue to the directory where the files are unzipped or to the installation CD. If this is an update to an earlier driver, perform the following steps:

1. Open Device Manager and right click (the first NetXen adapter) **NetXen Dual 10G Ethernet Adapter** → **Properties** → **Driver** → **Update Driver**. See Figure 5-4 on page 65.

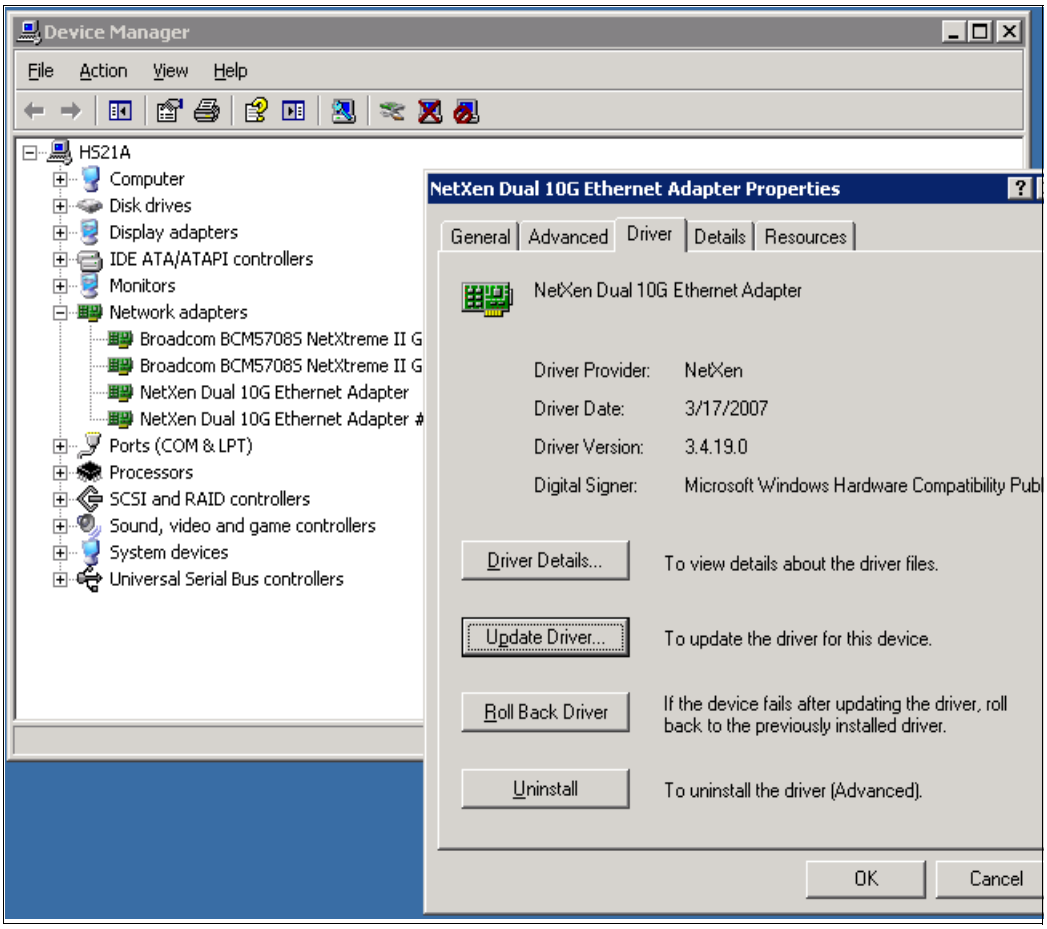


Figure 5-4 NetXen driver 10 Gb Ethernet Adapter properties

2. On the next window (Figure 5-5), select **No, not this time**. Click **Next**.

Figure 5-5 Hardware Update Wizard

3. Select **Install from a list or specific location (Advanced)** (see Figure 5-6**)**. Click **Next** to continue.

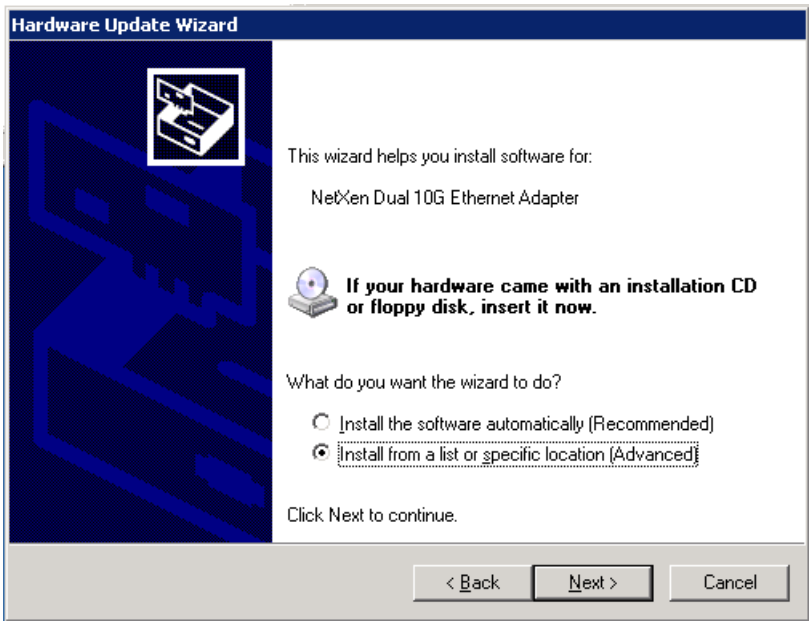


Figure 5-6 Hardware Update Wizard

4. On the next window (Figure 5-7), browse to the directory that contains the drivers or to the installation CD. Note that there may be two different versions of the driver, one in a subdirectory called Win2K3 and one in a subdirectory called SNP. The SNP (Scalable Networking Pack) version of the driver is required to use the TCP/IP Offload functionality of the NetXen card. This version of the driver may require a license key. In addition, in each of the subdirectories mentioned above there will be x64 and x86 directories, for 64- and 32-bit versions of Windows 2003. Click **Next** to continue once you have browsed to the directory containing the appropriate version of the driver.

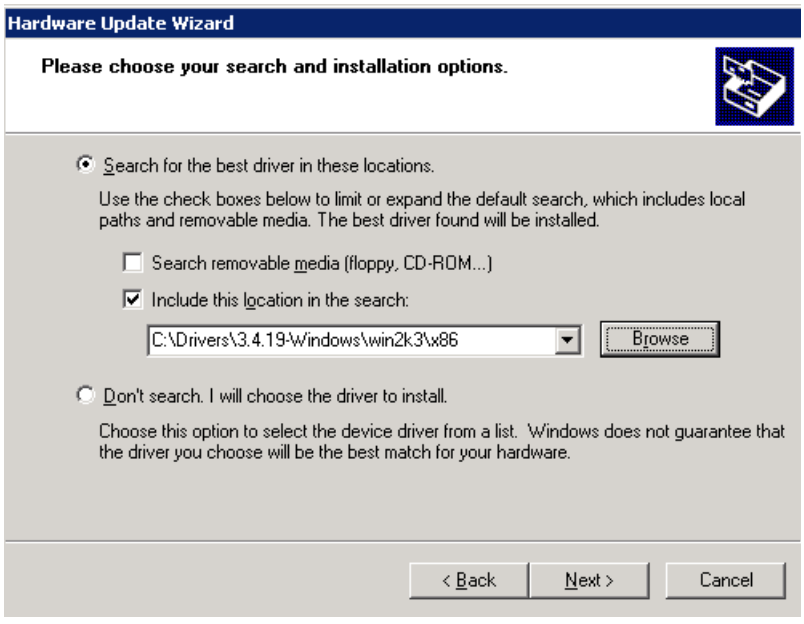


Figure 5-7 Search and installation options

Repeat these steps for the second NetXen 10 Gb Ethernet Expansion Card for IBM eServer BladeCenter.

5.5.3 Firmware

When the driver is installed, check if the firmware on the card is at the correct level. Use the instructions that are in the readme file.

1. Once the 3.4.216 driver is already loaded, identify the current Firmware version.
   a. Use the flash utility from the utils directory.
      i. nxflash -i NX_NIC –info (the NX_NIC is case sensitive)
   b. If not 3.4.216, update the firmware (DO NOT interrupt the flash update process until completed).
      i. nxflash -i NX_NIC –all
      ii. When the firmware update is complete, REBOOT the host.
      iii. See the Device Property sheet on Windows to use various setup options for the NIC.

Example 5-1 Screen of firmware update

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

C:\Drivers\3.4.216-Windows\utils>dir
 Volume in drive C has no label.
 Volume Serial Number is BCE8-DDA1

 Directory of C:\Drivers\3.4.216-Windows\utils

05/18/2007  04:46 PM    <DIR>          .
05/18/2007  04:46 PM    <DIR>          ..
05/16/2007  07:47 PM         4,194,304 imez_romimage
05/15/2007  08:45 PM           143,360 nxflash.exe
05/15/2007  08:45 PM           237,568 nxlic.exe
05/15/2007  08:45 PM            73,728 nxudiag.exe
05/16/2007  07:46 PM         4,194,304 xfp_ibm_romimage
               5 File(s)      8,843,264 bytes
               2 Dir(s)     636,215,296 bytes free

C:\Drivers\3.4.216-Windows\utils>nxflash -i NX_NIC --info
Supported flash
Board Type       : XGb IMEZ
Board Chip Rev   : C1
Serial Id        : MZ6BMK0049
Firmware version : 3.4.216
BIOS version     : 1.0.2
MAC Addr 0       : 00:0E:1E:00:12:00
MAC Addr 1       : 00:0E:1E:00:12:01
Subsystem ID     : 0x00044040

C:\Drivers\3.4.216-Windows\utils>nxflash -i NX_NIC --all
Supported flash
Currently in flash ::
Board Type       : XGb IMEZ
MAC Address 0    : 00:0E:1E:00:12:00
```

```
MAC Address 1    : 00:0E:1E:00:12:01
Serial Number    : MZ6BMK0049
Rom Image : imez_romimage
NIC binary romimage found in C:\Drivers\3.4.216-Windows\utils
Restoring License...
 100%  - DONE
No VPD area in romimage
Updating MD5...
Preparing backup...
 100%  - DONE
Backup file : "flashbackup.Fri-May-18-16-58-03-2007" - completed successfully.
WARNING: This is a very sensitive operation.
Do not interrupt until operation is complete.
Taking NetXen NIC offline...
Erasing flash...



WARNING: This is a very sensitive operation.
Do not interrupt until operation is complete.
Taking NetXen NIC offline...
Erasing flash...
Starting flash process...
 007%

Taking NetXen NIC offline...
Erasing flash...
Starting flash process...
 100%  - DONE
Flashing completed successfully.
Reboot system or reload driver for firmware to take effect
NetXen NIC is online now...
--------------------------------------------------------------------------------
```

5.5.4 Teaming

NIC teaming in Windows is provided in the Intermediate Driver (IM) which is supplied by NetXen, the manufacturer of the NetXen 10 Gb Ethernet Expansion Card for IBM eServer BladeCenter, or can alternately be provided by the existing Broadcom BASP teaming software provided by Broadcom for the on-board Broadcom NICs. The Broadcom teaming driver supports all the various NICs supported in the BladeCenter (Broadcom, Intel®, NetXen), but only provides full functionality like 802.1q VLAN tagging, on the Broadcom NICs. To get the full functionality of the NetXen 10 Gb Ethernet Expansion Card for IBM eServer BladeCenter the NetXen IM driver should be used.

This driver also provides teaming for the onboard Broadcom adapter, this includes advanced functions like 802.1Q.

The Broadcom BASP driver and the NetXen IM driver cannot be used at the same time on the same system.

5.5.5 Installing the NetXen teaming driver

To install the NetXen teaming software, perform the following steps:

1. Open the Network Connections folder and right click on one of the installed NetXen Dual 10G Ethernet Adapters. Click **Properties** → **Install** → **Protocol** → **Add**. See Figure 5-8.

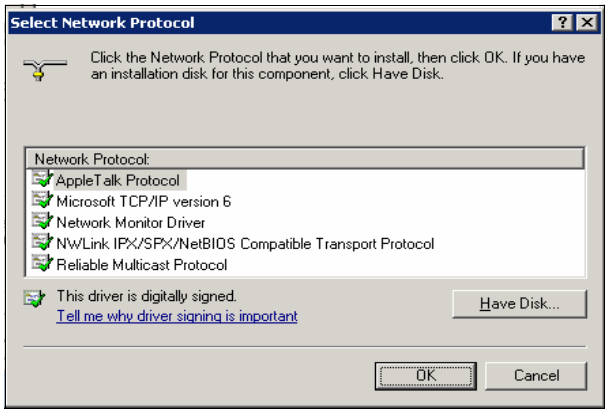


Figure 5-8 Select Network Protocol window

2. Click **Have Disk**. You will see a window similar to Figure 5-9.

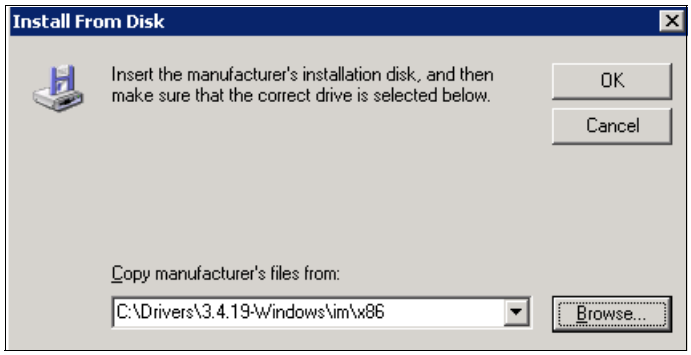


Figure 5-9 Install From Disk window

3. Browse to the directory or CD containing the NetXen IM drivers.
4. Highlight nxteam.inf driver (Figure 5-10 on page 70). Click **Open** to continue.

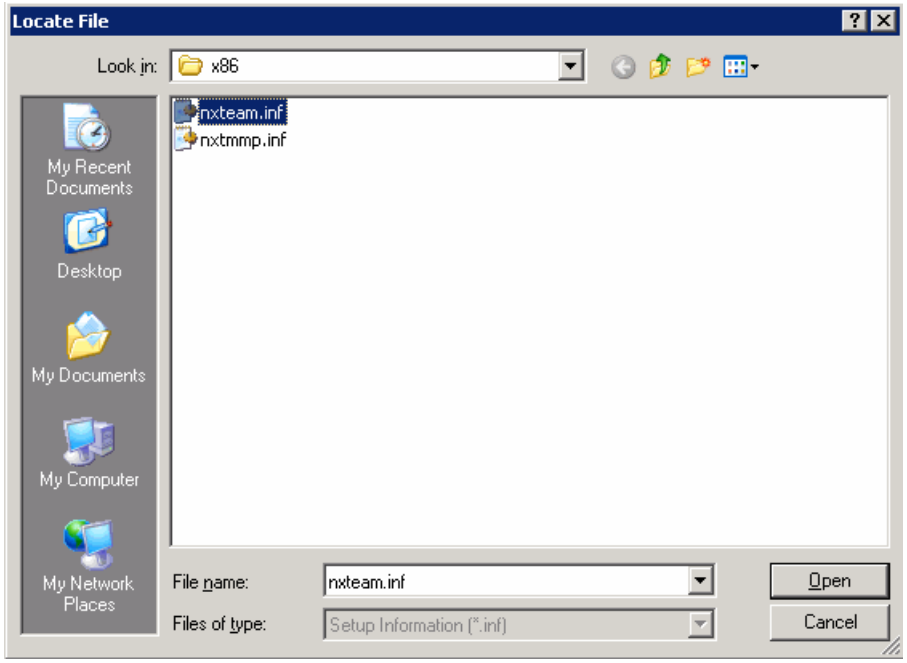


Figure 5-10 Locate file window

5. Highlight NetXen Network Configuration Utility 7 (Figure 5-11) and click **OK**.

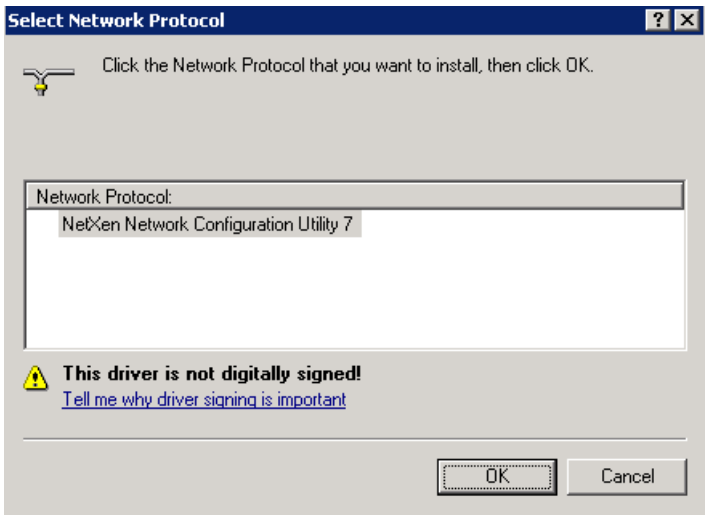


Figure 5-11 Select Network Protocol window

6. The next window (Figure 5-12 on page 71) will contain a warning about Windows Logo Testing. Click **Continue Anyway** to ignore this.

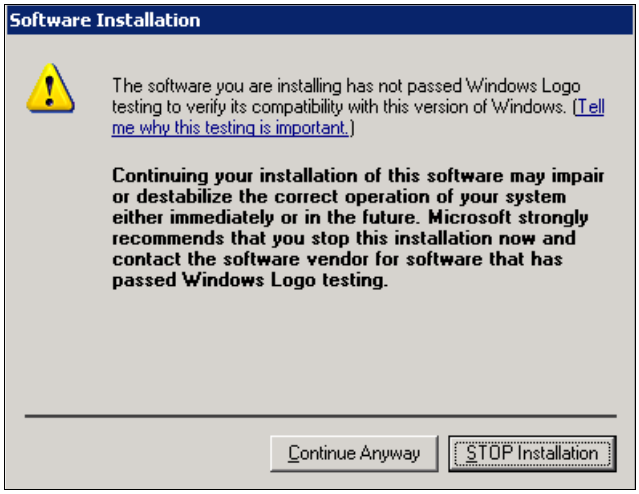


Figure 5-12 Software Installation window

7. Under the General tab (Figure 5-13), the NetXen Network Configuration Utility for teaming is now added to the list of Protocols.

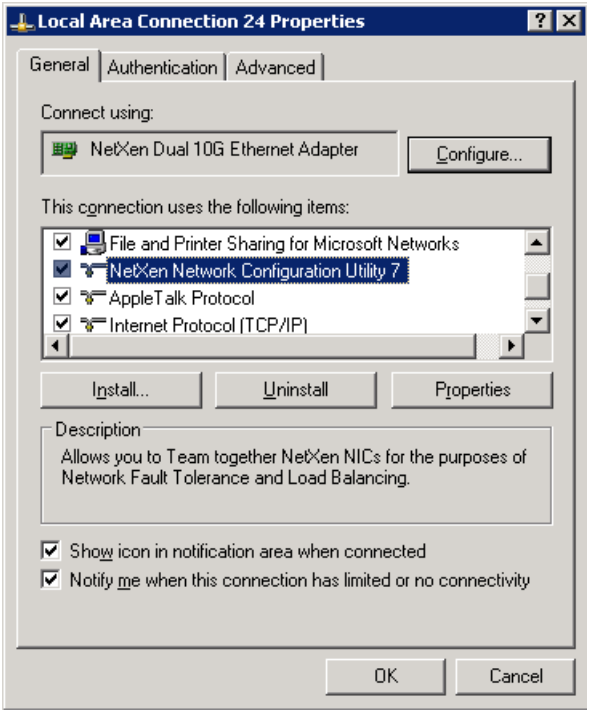


Figure 5-13 General tab window

5.5.6 Configuring the NetXen IM driver

Once the NetXen IM driver is installed, teaming and other advanced functions can be configured (see Figure 5-14 on page 72).

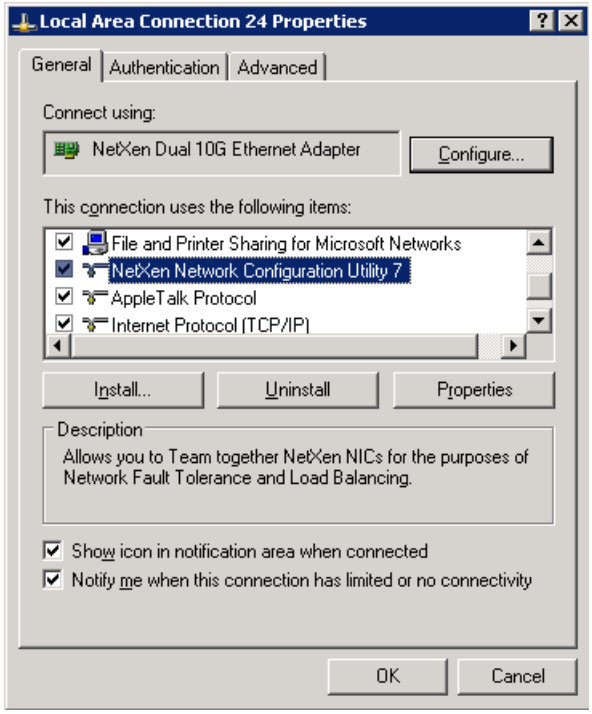


Figure 5-14 General tab window

1. Open the property screen of one of the installed NetXen 10 Gb adapters in the Network Connections window by right clicking the icon, and highlight **NetXen Network Configuration Utility 7**. Click **Properties** to open the Configuration Utility. Or double-click the NetXen Network Configuration Utility 7 icon that has been placed in the Windows Taskbar. Either action opens the Configuration Utility window shown in Figure 5-15 on page 73.
2. Upon first opening of the Configuration Utility a NetXen icon is added to the Windows Taskbar. This can be removed by clicking the checkbox for the Icon in the NetXen Network Configuration Utility.

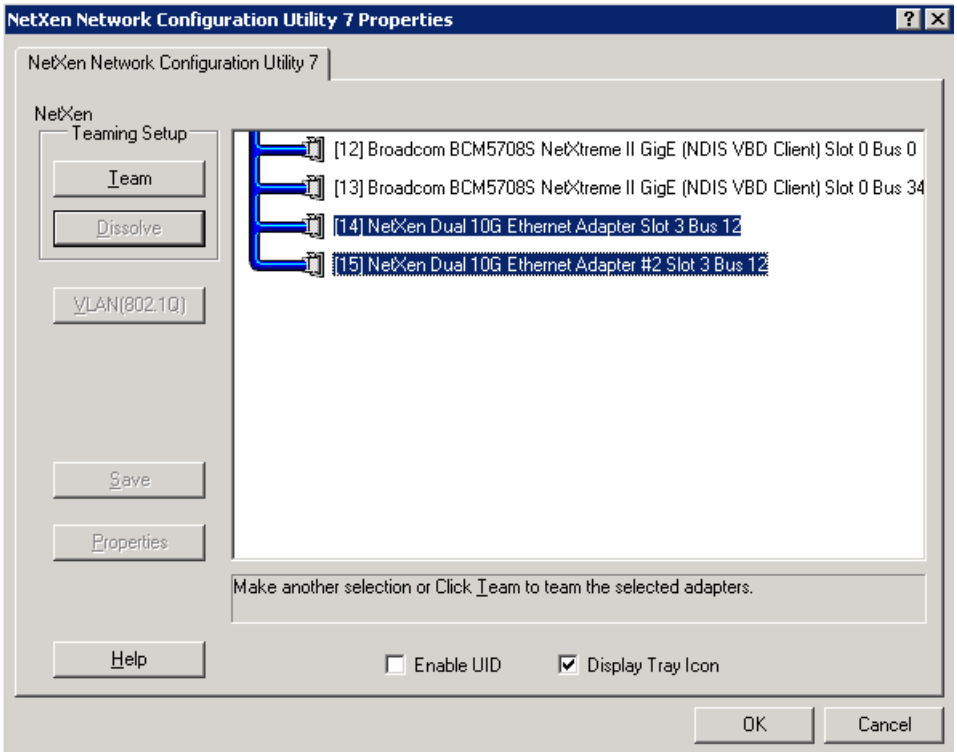


Figure 5-15 NetXen Teaming Setup window, note the Display Tray Icon

3. Select the adapters that make up the team. For our example, NetXen Dual 10G Ethernet Adapter Slot 3 Bus 12 and NetXen Dual 10G Ethernet Adapter #2 Slot 3 Bus 12.

   The NetXen Network Configuration Utility shows all the network adapters in the system. Two Broadcom and two NetXen cards on a single width blade such as the HS21 and LS21, and four Broadcom and two NetXen on a double width blade such as the LS41. The utility can be used to team either the 1 Gb Broadcom or the 10 Gb NetXen adapters. It is not possible to create a mixed team.

4. To create a new team, highlight both NetXen adapters and click **Team**. This opens the following window:

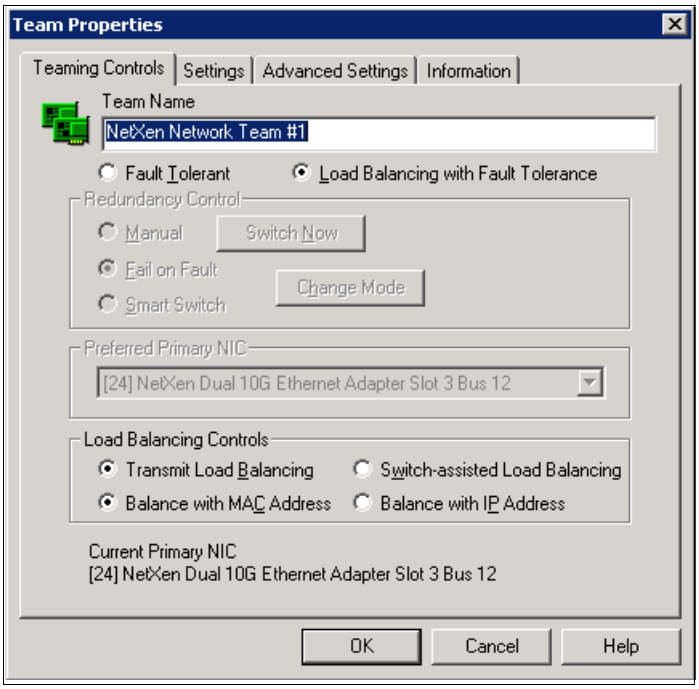


Figure 5-16 Team Properties window

The default preselected teaming method setting in the NetXen IM driver is Load Balancing with Fault Tolerance. The default settings for the Load Balancing Controls are Transmit Load Balancing and Balance with MAC address. These settings are preselected on opening the Team Properties window.

Note: In the current version of the NetXen IM driver software, the team has to be created with the initial default settings. After the team is created successfully, it needs to be changed to Fault Tolerant. (Fault Tolerant is the only recommended teaming method between the blades and I/O modules.)

Fault Tolerance Options (Redundancy Control)

The NetXen adapter provides three fault tolerant options:

Manual: This setting allows change from a Primary NIC to a secondary NIC (only when Switch Now is clicked).

Fail on Fault: This setting automatically switches from a primary NIC to a secondary NIC when the primary NIC fails.

Smart Switch: This setting lets a member of a team be selected as the preferred Primary Smart Switch NIC. As long as this NIC is operational, it is always the active NIC. If the NIC fails and it is eventually restored or replaced, it automatically resumes its status as the active NIC. Smart Switch is the recommended choice for fault tolerance.

5. Click **OK** to accept the default settings and the teaming configuration is created. This can take a few minutes. Click **Continue Anyway** (Figure 5-17 on page 75) if the Microsoft Windows Logo message displays.

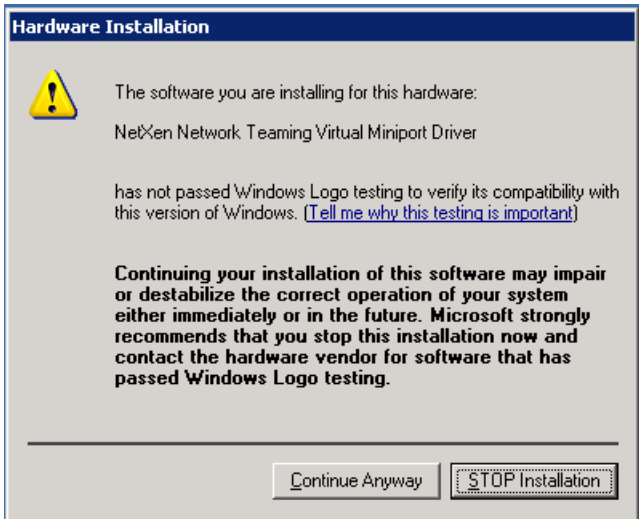


Figure 5-17 Hardware Installation window

6. When the teaming configuration has finished, the following message (Figure 5-18) will display.

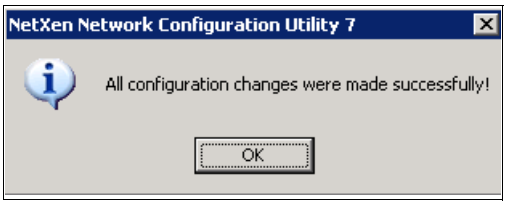


Figure 5-18 NetXen Network Configuration Utility 7 window

7. Back in the NetXen Network Configuration Utility (Figure 5-19 on page 76), the team now appears with a green square showing the existence of a team.

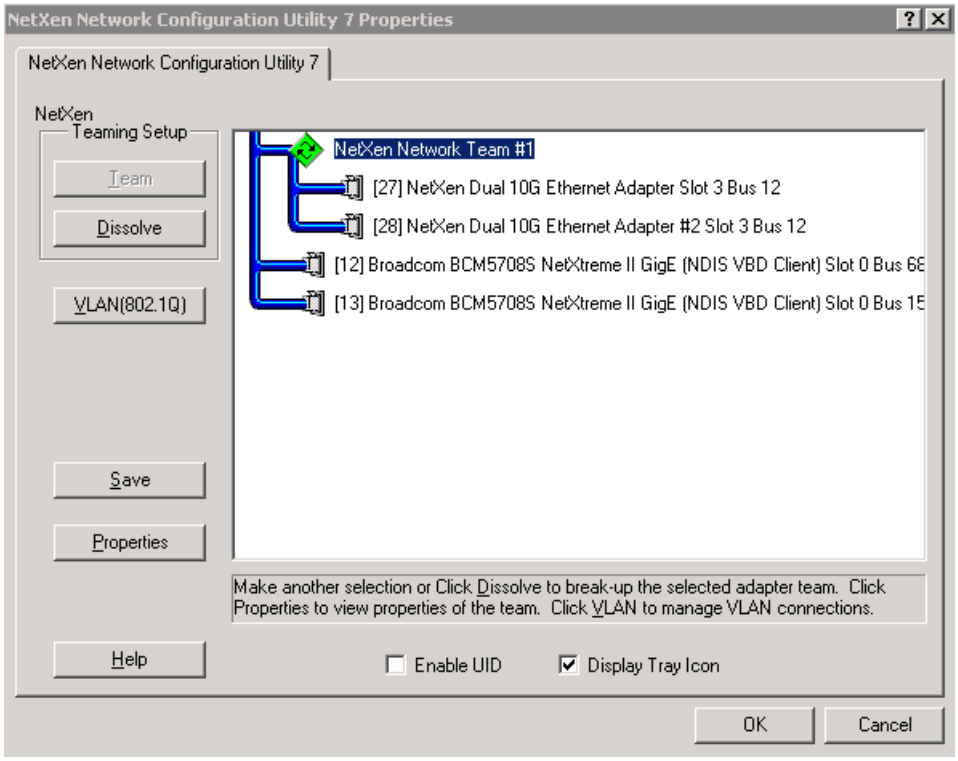


Figure 5-19 NetXen Network Configuration Utility 7 window

5.5.7 Removing the NetXen IM Driver

Make sure that no teams exist in the NetXen Utility. Also remove the tick from the display tray icon in the NetXen Network Configuration Utility. If you do not remove the display tray icon the icon will stay in existence even after uninstalling the driver.

The IM service for nxteam.exe is located in C:\windows\system32.

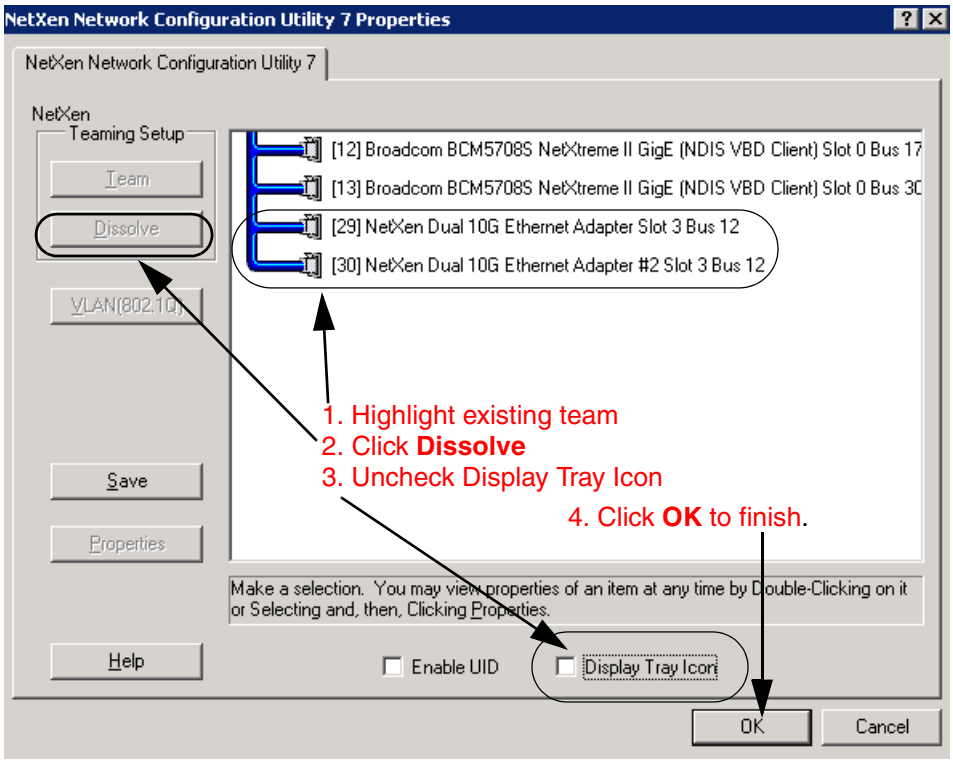


Figure 5-20 Removing the existing configuration

1. From the NetXen Network Configuration Utility 7 window, highlight the NetXen Dual 10G Ethernet Adapters in Figure 5-20. Click **Dissolve**. Uncheck the Display Tray Icon. Click **OK.** Open the Properties page of one of the NetXen adapters. Highlight the NetXen Network Configuration Utility 7 in the Protocol list and click **Uninstall**. See Figure 5-21.

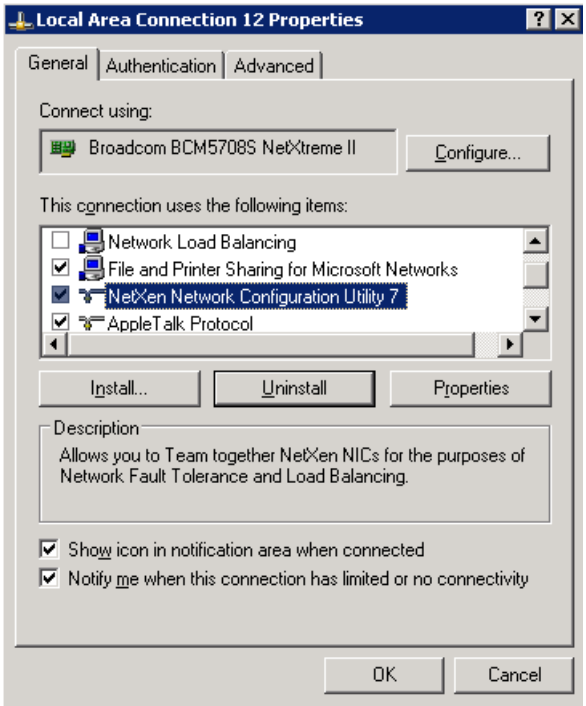


Figure 5-21 General tab window

2. Answer Yes to the question if you want to Uninstall the component from all Network Connections, and to the following window that tells you need to restart the computer.

5.5.8 Alternate configuration: Using Broadcom software

The NetXen 10 Gb Ethernet Expansion Card for IBM eServer BladeCenter can also be teamed using the existing Broadcom teaming drivers that are used for the onboard Broadcom NICs. Be aware that the BASP software does not support VLAN tagging or other advanced options on non Broadcom NICs. Ensure that the NetXen adapter driver is installed, but that the NetXen IM (Teaming) driver is not loaded. If the NetXen IM driver was previously installed it needs to be completely removed by following the steps in the previous section; the two drivers cannot coexist.

The BASP driver is a generic driver, (for example, not specific to BladeCenter) that provides four different teaming methods:

- Smart Load Balancing (TM and Failover)
- Link Aggregation (802.3ad)
- Generic Trunking (FEC/GEC)/802.3ad-Draft Static
- SLB (Auto-Fallback Disable)

Only option 1 and 4 (Figure 5-22) are supported on the blades. It is not possible to use option 2 and 3 as in this configuration; both links should be connected to the same switch. This is not possible within the hardware configuration of the BladeCenter.

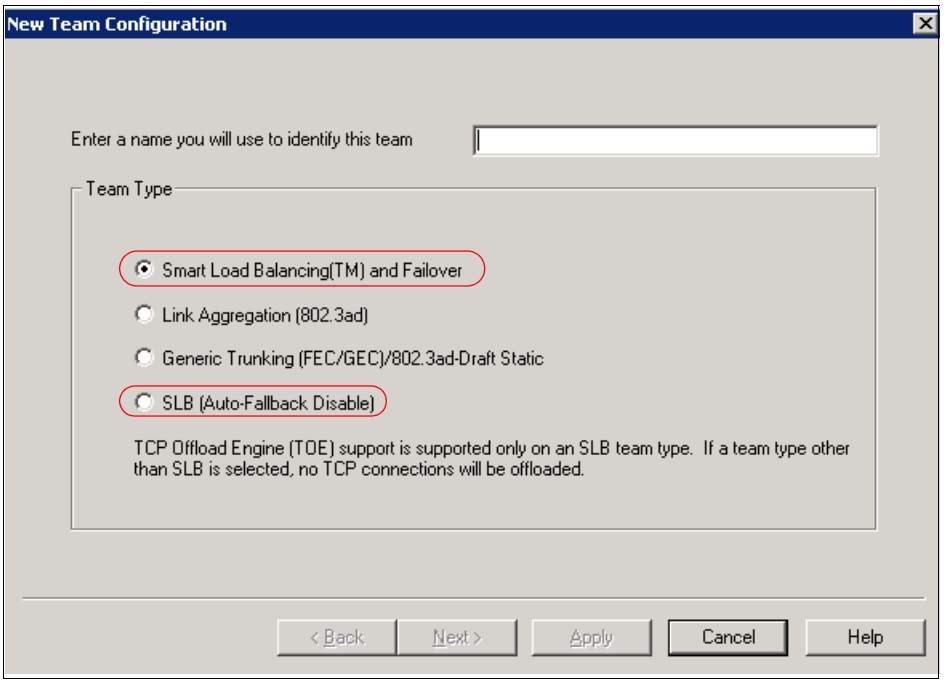


Figure 5-22 Option 1 and 4 only

5.5.9 Installing the Broadcom BASP teaming driver

Create a Broadcom driver CD or copy the Broadcom drivers to a directory on the blade. After the blade has rebooted following the installation of the NetXen card or after removal of the NetXen IM drivers, enter the Broadcom CD or click launch.exe in the directory where you have copied the Broadcom software.

The Broadcom NetXtreme Software installer will launch automatically when the CD is inserted. Click **Management Applications** (Figure 5-23) and the installshield Wizard program will be launched.

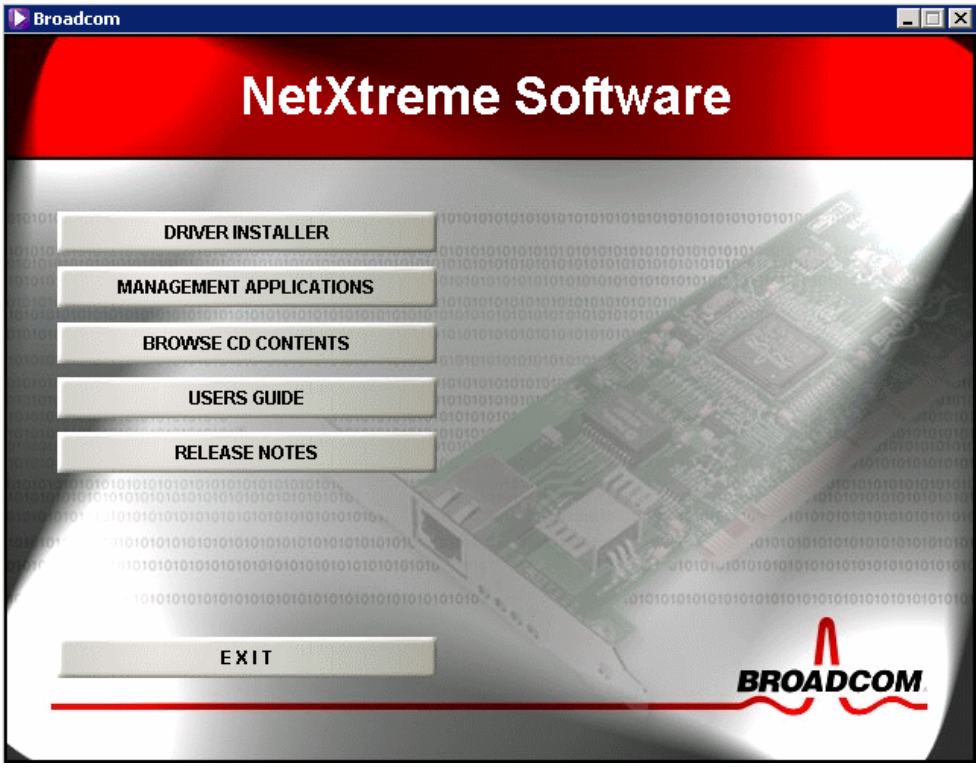


Figure 5-23 The Broadcom NetXtreme Software Installer

1. Click **Next** on the Broadcom Welcome window

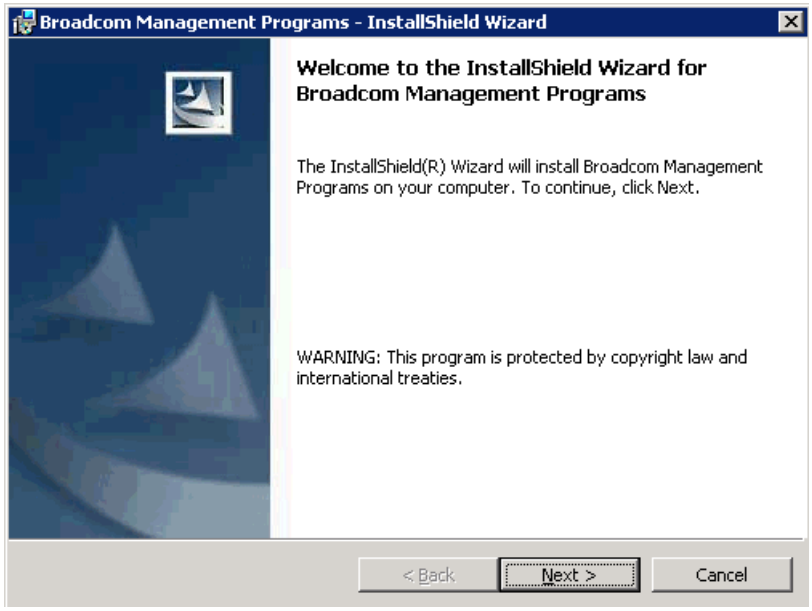


Figure 5-24 the Broadcom Welcome window

2. Accept the License agreement (Figure 5-25) and click **Next**.

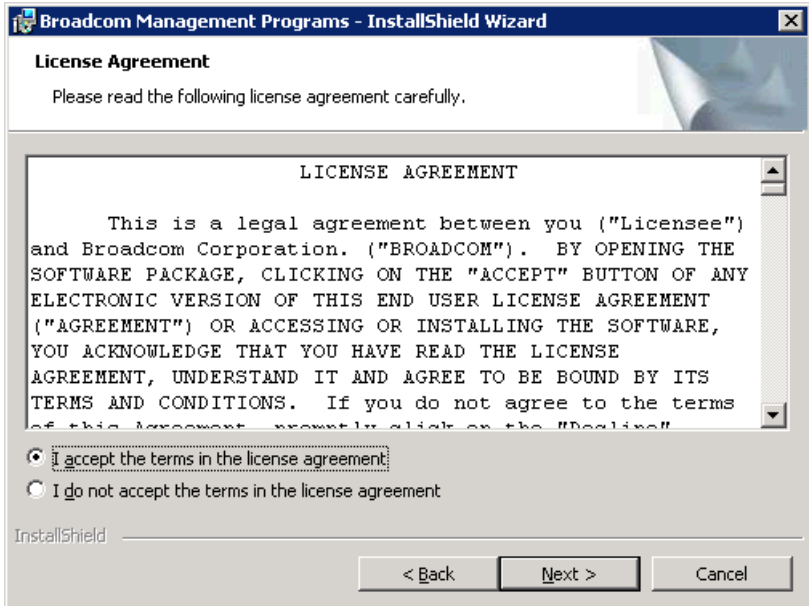


Figure 5-25 The License agreement

3. On the Custom Setup window (Figure 5-26), the Broadcom Advanced Control Suite (BACS) graphical user interface and the Broadcom Advanced Server Program NDIS intermediate driver (BASP driver) are preselected. At this stage, do not install the SNMP or CIM software. Click **Next** to continue.

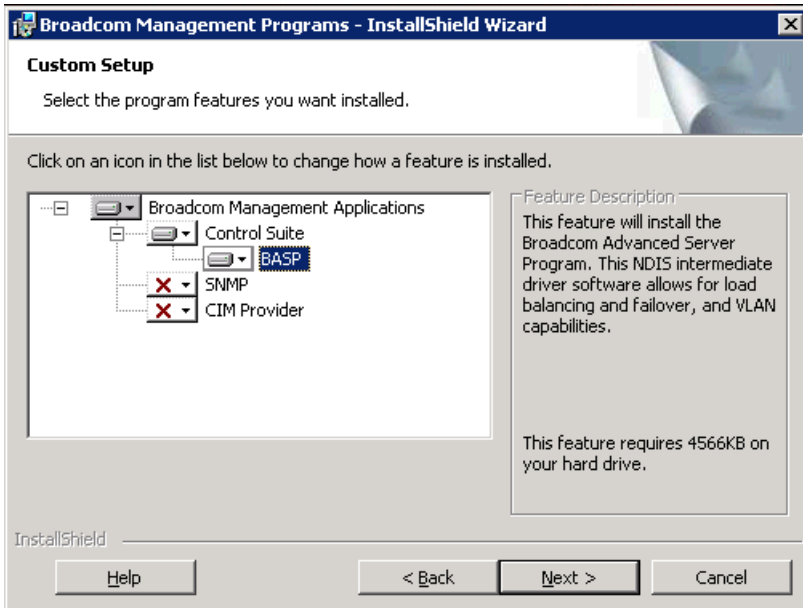


Figure 5-26 The custom setup window

4. Click **Install** to continue the installation. See Figure 5-27.

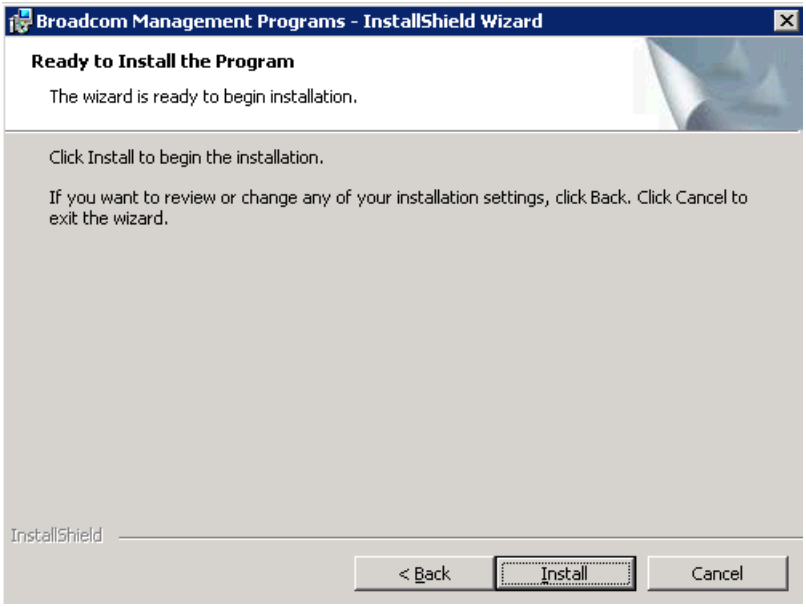


Figure 5-27 Ready to Install

5. The files are installed successfully (Figure 5-28). There is now a BASC control suite icon in the taskbar. Click **OK** to continue.

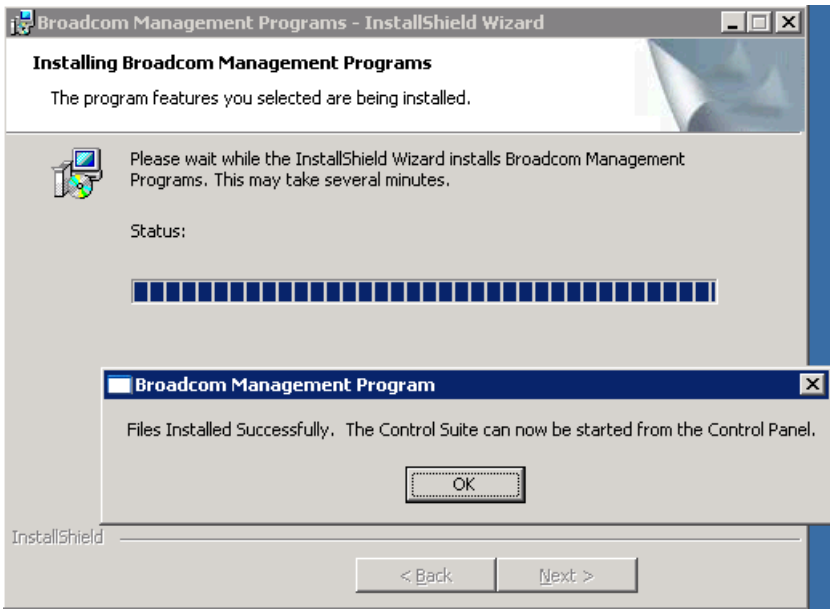


Figure 5-28 Success window

6. Click **Finish** (Figure 5-29 on page 82). This completes the installation of the Broadcom software. Click **Exit** to close the Broadcom NetXtreme Software installation utility.

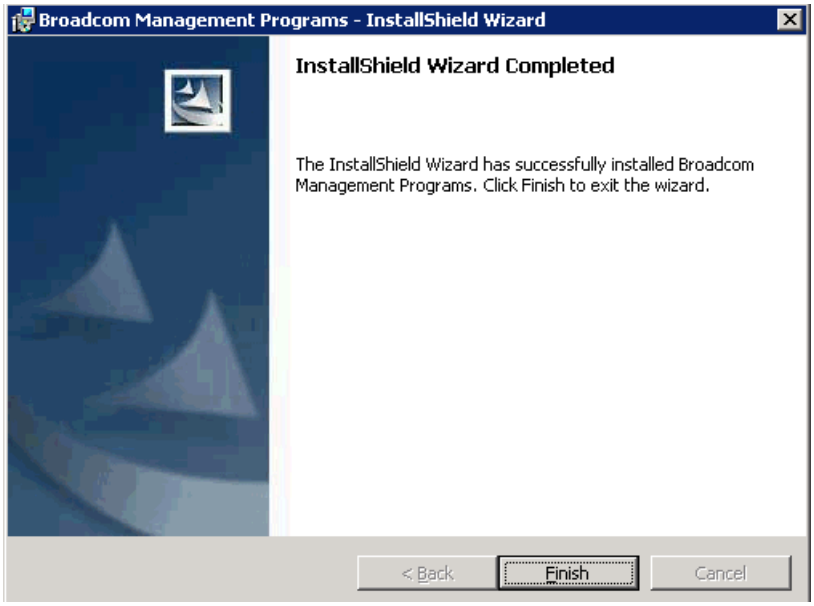


Figure 5-29 Installation complete

The next section will cover configuration of the BASP teaming driver using the BASC utility.

5.5.10 Configuring teaming with the Broadcom driver

This section demonstrates how to configure teaming with the Broadcom driver.

1. Double-click the green magnifying icon in the taskbar to start the Broadcom Advanced Control Suite 2 (BACS).

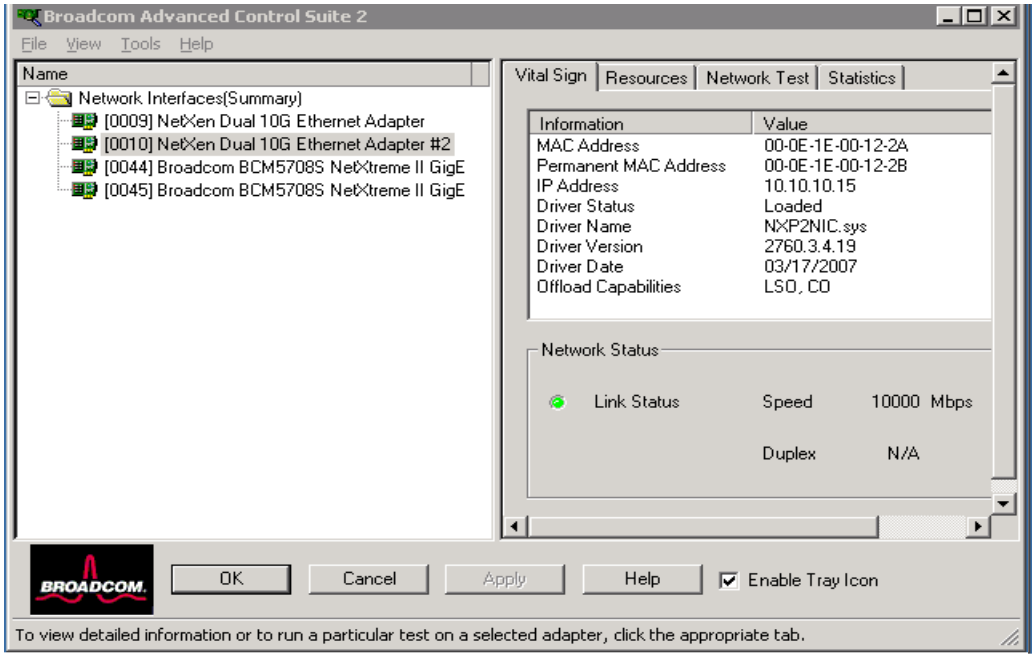


Figure 5-30 Broadcom Advanced Control Suite

2. The Vital Sign tab (Figure 5-30 on page 82) shows layer 2 and layer 3 addressing information, driver version, and capability information. Note that TOE is not supported on the NetXen adapter using the Broadcom driver, but Large Send Offload (LSO) and Checksum Offload (CO) are supported.
3. Pull down the tools menu and select create a team. The first time, this will open in wizard mode. See Figure 5-31.

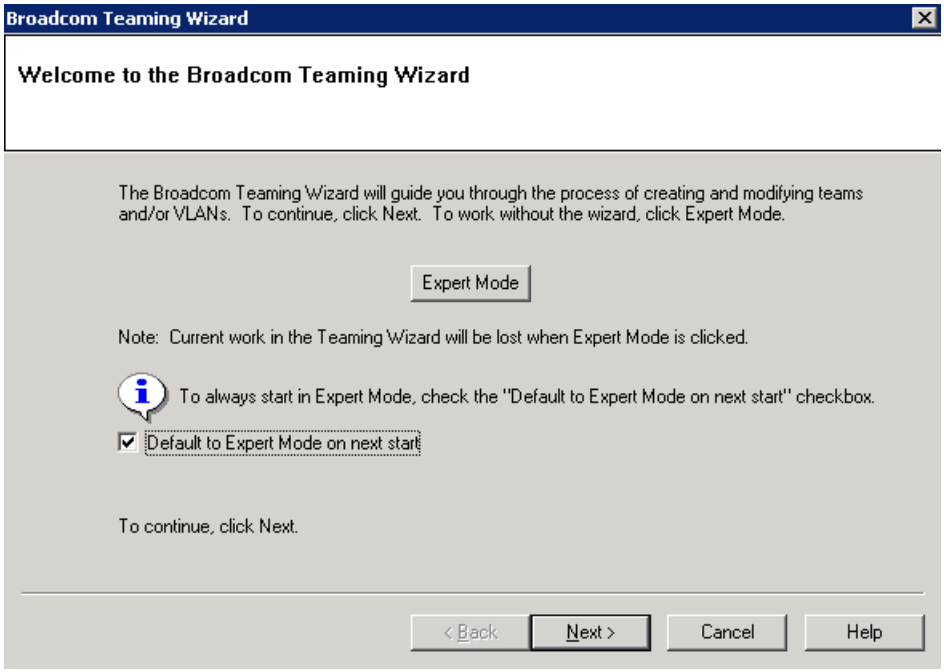


Figure 5-31 Expert or Wizard

4. Of course, we all like to be known as experts, and therefore we do not need Wizards. So check the checkbox for **Default to expert mode on next start** and click the **Expert mode** button.

   The default teaming mode can be changed back to Wizard by pulling down the Tools menu and selecting Customize.

   Clicking the Expert mode button brings up the new team configuration window. See Figure 5-32 on page 84.

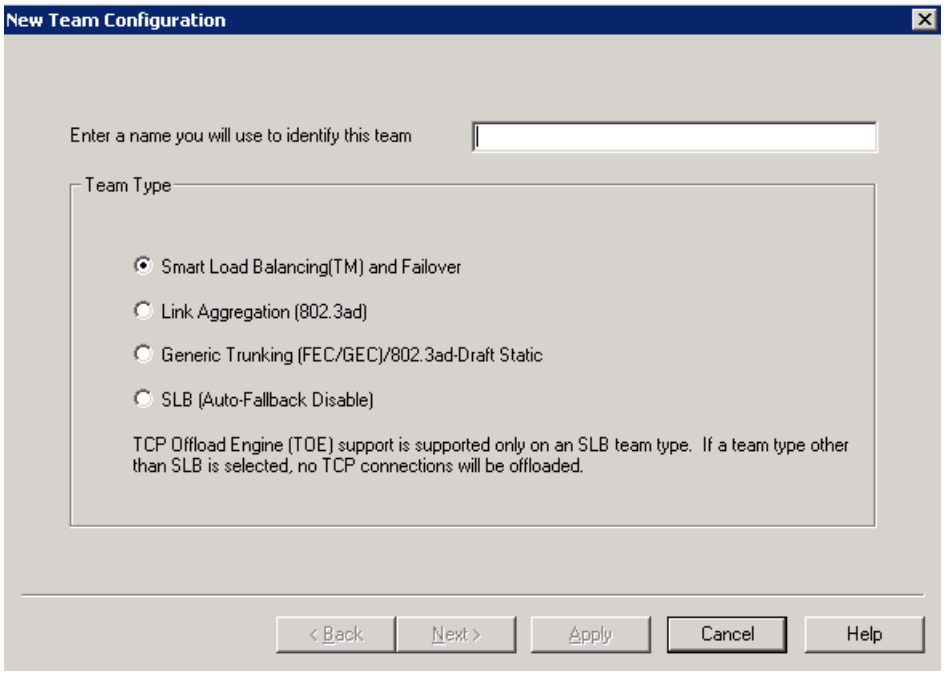


Figure 5-32 Team configuration window

5. Choose a name (something like *Team-1*) and choose your team type. Only option 1 and 4 are supported on the BladeCenter because both Link Aggregation and Generic Trunking only work when connected to the same physical switch.

   The difference between the two supported options:

   – Smart Load Balancing(TM) and Failover
   – SLB (Auto-Fallback Disable)

   With the first option, the team will fall back to the original primary adapter when the link is restored. With Auto-Fallback Disable, the new primary adapter status will remain on the second adapter.

6. On the next page (see Figure 5-33 on page 85), select the adapters that will be part of the team. Select one adapter as a Load Balance member and another adapter as a Standby member. Again, this is dictated by the hardware layout of the BladeCenter; a team needs to be in active/standby mode to work properly.

   An active/active configuration will also work, but this will cause issues with forwarding tables on the switches, active/passive is the only recommended method.

   If this causes concerns about throughput performance, one solution is to set-up the NICs in half of the blades as active/standby and the other half as standby/active so the load will be equally balanced between the Nortel 10 Gb Ethernet Switch Modules.

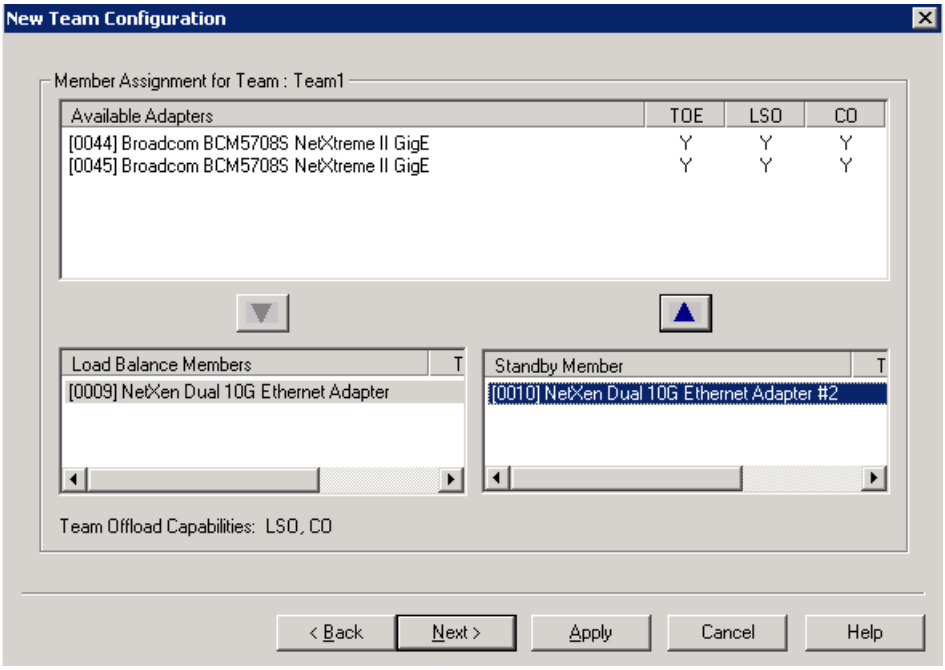


Figure 5-33 Choose your adapters

7. Place one adapter in the Load-Balance member group and one in the Standby members group. A warning will display (Figure 5-34) stating that the NetXen Dual 10G Ethernet Adapter is not fully certified. Click **Yes** to continue.

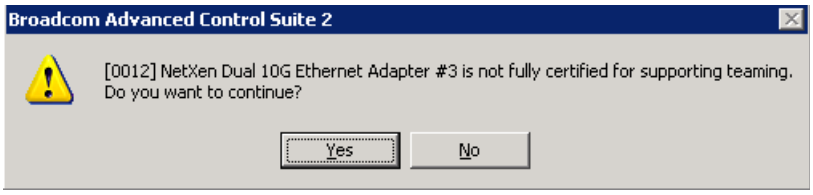


Figure 5-34 Ignore this warning

When both adapters are assigned to their group, the window should look similar to Figure 5-35 on page 86.

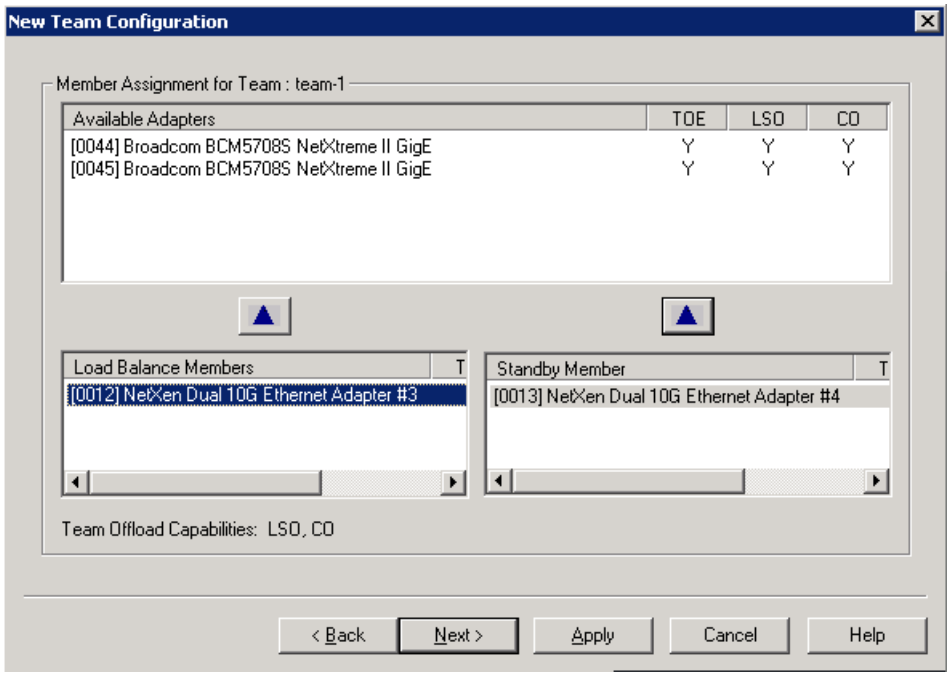


Figure 5-35 Member Assignment for Team: Team 1 window

8. Click **Apply** to configure this team; a warning will display (Figure 5-36) stating that applying the changes will temporarily interrupt the network connection. Click **Yes** to continue.

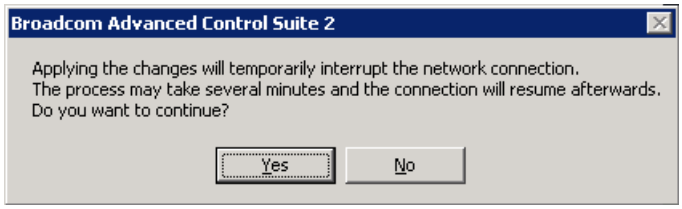


Figure 5-36 Dialog box

9. Click **OK** to set IP address reminder (Figure 5-37).

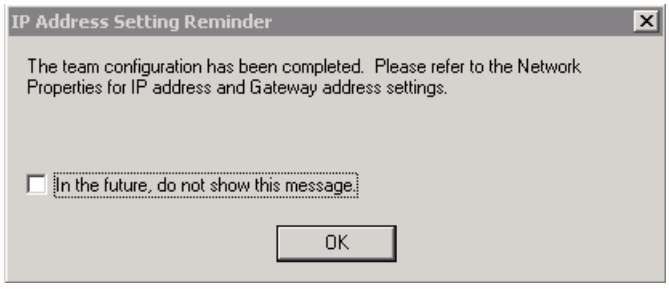


Figure 5-37 Reminder to set the IP address

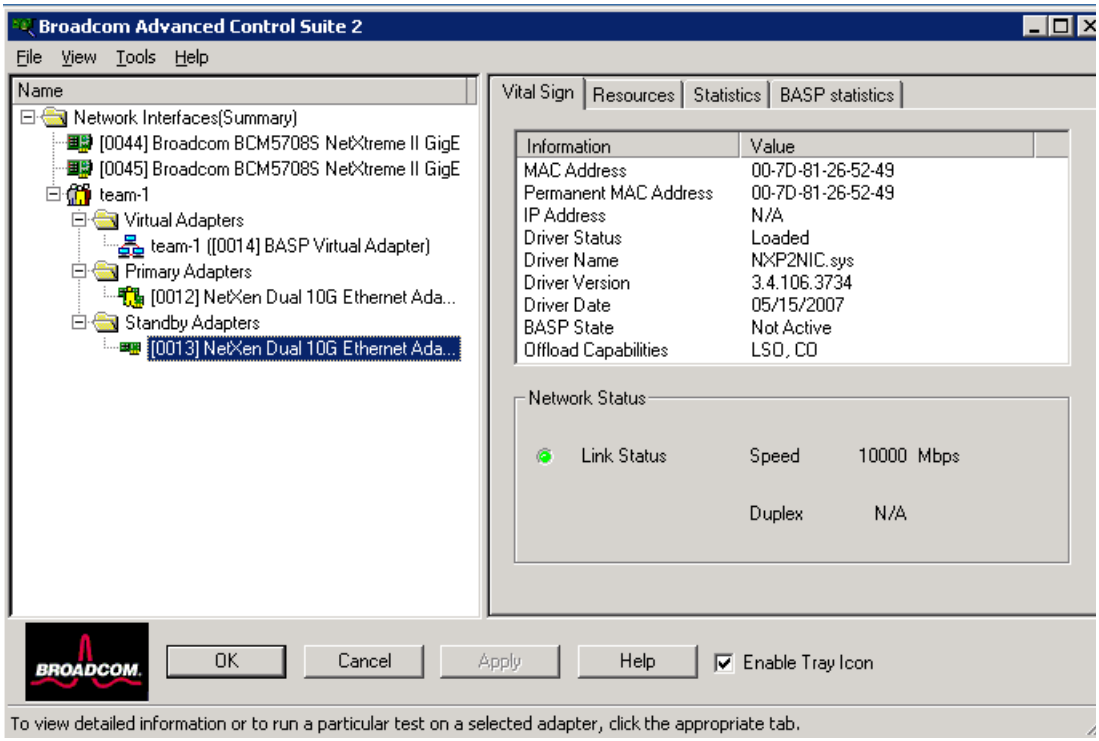


Figure 5-38 The configured team

The last screen shows the configured team with a BASP virtual adapter consisting of the two NetXen adapters. Additionally the two Broadcom adapters can be teamed from this screen. When you highlight the Virtual Adapters Icon in the Name screen, there are some more options that can be configured: LiveLink, Delete Team, Configure Team, and AddVlan. The AddVlan option, which configures 802.1q tagging, is only supported on the Broadcom adapters.

5.6 Installing the NetXen drivers on Linux servers

The process for installing the drivers on Linux systems is the same for both Red Hat and SUSE, therefore, the procedures will be covered in this section. Installation and configuration of the Linux Sockets Acceleration feature is the same for both Linux distributions and will be covered in a common section after the bonding sections. NIC bonding is slightly different between the two distributions, therefore, those steps will be described in the respective distribution's section.

The NetXen driver is supplied by IBM through the Document IDs below that link to the NetXen Web site. Registration is needed before download. Drivers for Linux are available in 32- and 64-bit versions. Power Linux drivers are in development.

The driver package contains:

- NIC Linux Driver
- TOE Linux Driver
- Firmware
- Diagnostics
- Tools
- Documentation in PDF format.

The GA level initial release version for Linux is: Version 3.4.216.

5.6.1 Installing the NetXen NIC Driver on Linux

Note: Install the Linux kernel sources/headers to compile the source code driver. Some Linux OS distributions might not install the kernel source/headers by default during installation. Add the kernel sources/header files from the Linux Operating System Distribution source. The operator must also have installed the compiler tools in order to compile the source distribution.

Note: The filenames of the tarball includes the version information, for example, 3.4.216.tar.gz.nx. The installation instructions are included with the tarball substitute *<version>* when referring to the filename.

The following steps will compile the NetXen driver on the Linux distribution:

1. Copy the tarball to a working directory.
2. Un-tar the driver package.

   ```
   tar xvzf 3.4.216.tar.gz.nx
   ```

3. Build and install the driver module.

   ```
   cd 3.4.216
   make
   make install
   ```

4. Load the driver module.

   ```
   insmod ./driver/nx_nic.ko
   ```

5. Bring up and verify all the Ethernet interfaces.

   ```
   ifconfig eth[x] <ipaddress> up
   ```

6. Check the status of the driver module loaded; value should be 2.

   ```
   lsmod | grep nx
   ```

7. Check for system messages related to the card.

   ```
   dmesg
   ```

8. Test the Ethernet interface using ping.

5.6.2 Updating the NetXen firmware on Linux servers

The following procedure assumes that the user has already installed the NetXen card and drivers, that a new version of the drivers and firmware has been released, and the user wants to install the new firmware and drivers on the NetXen 10 Gb card.

Note: During the firmware update process, do not reboot or stop the firmware process, this might corrupt the firmware.

There are checks built into the flash utility, and during the firmware update, the MD5 recovery step can identify incorrect firmware and stop the update process.

The flash tool will not start the firmware update process if the Ethernet interface is not up.

In the event of running the diagnostics, if it is found that there are identifiable errors in the log, the flash utility's recover option can be used to revert back to the previous version and restore the firmware image.

If an existing driver is already loaded, identify the current firmware and driver version using the flash utility from the drivers bin subdirectory:

```
./nxflash -i eth[x] --info
ethtool i eth[x]
```

Note: If you do not use the ./ prior to the nxflash command, it is possible you will execute an older version and you will experience unpredictable results.

Remove the old NIC driver:

```
rm /lib/modules/<kernelversion>/kernel/drivers/net/nx_nic.ko
```

Note: Kernel version can be determined by typing: uname -r

Install the new driver.

The steps required to install the NetXen NIC drivers are as follows:

1. Un-tar the driver package.

```
tar xvzf <version>.tar.gz.nx
```

2. Build and install the driver module.

```
cd <version>
make
make install
```

3. Load the driver module.

```
insmod ./driver/nx_nic.ko
```

4. Bring up and verify all the Ethernet interfaces.

```
ifconfig eth[x] <ipaddress> up
```

5. Check the status of the driver module loaded; value should be 2.

```
lsmod | grep nx
```

6. Check for system messages related to the card.

```
dmesg
```

7. Test the Ethernet interface using ping.

To update the firmware, use the flash utility, nxflash.

Note: Do not interrupt the flash update process until completed.

```
nxflash -i NX_NIC --all
```

When the firmware update is complete, reboot the host.

5.7 Linux Sockets Acceleration

This section discusses the capabilities of Linux Sockets Acceleration. Linux Sockets Acceleration provides the following:

- LSA reduces host CPU utilization, providing datacenters with the opportunity for better power efficiency and scalability.
- LSA supports both kernel- and user-mode applications.
- Yet LSA requires:
  - No changes to the kernel or stack
  - No recompile of the kernel (because LSA is a simple loadable module like any other driver)
  - No changes to any APIs
- LSA uses Selective Acceleration technology to let users decide what is accelerated and what is not. This technology puts control into users' hands—a hallmark of Linux and the open source movement.

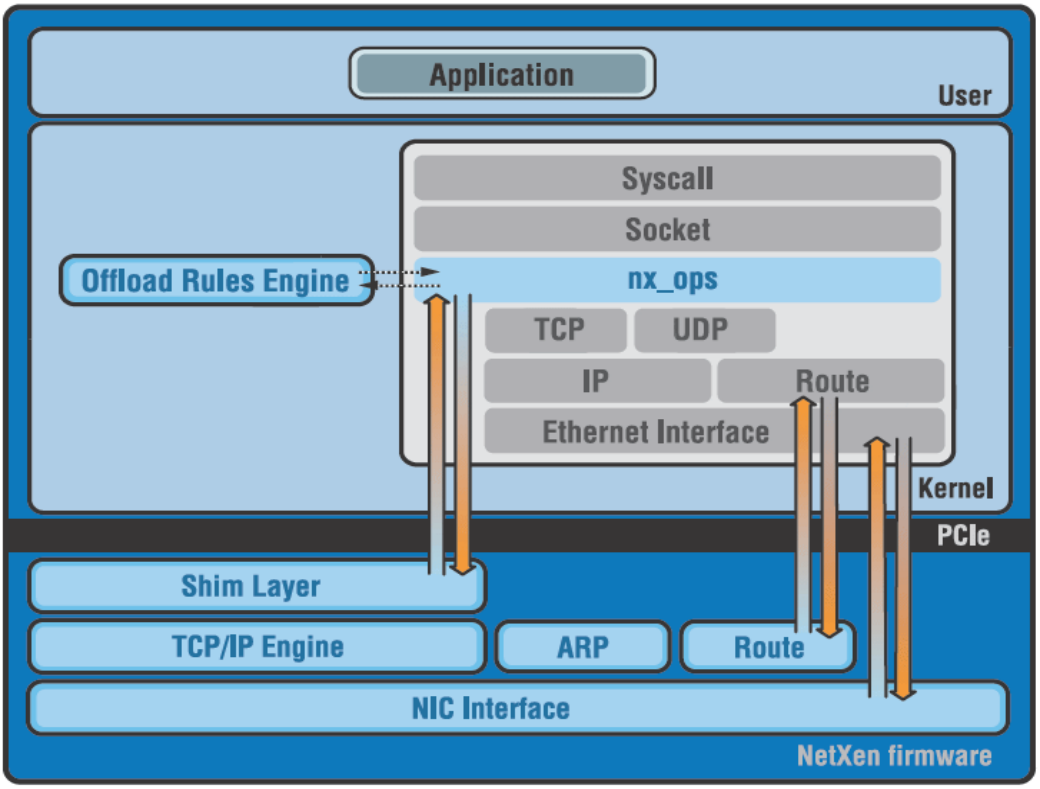


Figure 5-39 Linux Sockets Acceleration architecture diagram

5.7.1 LSA Driver Installation

Linux Driver Ordering: Load the Linux NIC driver (nx_nic) first, followed by the intercept driver (nx_intercept.ko), then the lsa driver (nx_lsa). The intercept driver must be running before loading and using the Linux LSA driver.

Note: The filenames of the tarball includes the version information, for example, 3.4.216-lsa.tar.gz.nx. The installation instructions are included with the tarball substitute <version> when referring to the filename.

First, ensure that the NIC driver is installed and running as described in Section 5.6, “Installing the NetXen drivers on Linux servers” on page 87.

If a previous version of the LSA driver is installed on the server, remove the driver by following the directions in section 5.7.3, “Uninstalling the LSA driver” on page 92.

1. Copy the LSA driver tarball to a working directory.
2. Build the intercept and LSA driver, see Example 5-2.

Example 5-2 tar 3.4.216-lsa.tar.gz.nx file

```
tar xzf 3.4.216-lsa.tar.gz.nx
cd 3.4.216-lsa
make
make install
```

The install script will copy the drivers to the /lib/modules/<Kernel Version>/kernel/drivers/net/ directory.

3. Load the Intercept driver.

   $ insmod ./driver/nx_intercept.ko

4. Load the lsa driver.

   $ insmod ./driver/nx_lsa

5.7.2 LSA offload configuration

Once the drivers are installed, the offload configuration is performed with the command-line utility *nxoffload*. The syntax for this command is shown in Example 5-3.

Example 5-3 nxoffload command syntax

```
$ nxoffload
usage: nxoffload <options>
options: -p   Port number to be offloaded.
         -n   Application name to be offloaded.
         -t   TCP tuple to be offloaded.
              Format: local IP, local port, remote IP, remote port.
              example:
                nxoffload -a -t'0, 0, 0, 5001'.
                nxoffload -a -t'0, 2000, 0, 5001'.
         -a   Add an offload rule.
         -r   Remove an offload rule.
         -z   Enable zero copy.
         -s   List all ports and applications offloaded.
```

You can check the offload statistics by reading the file "/proc/net/nx_nic/lsa_x/stats" where x is the NetXen card number.

5.7.3 Uninstalling the LSA driver

To remove the driver, first bring down the interface, and then use rmmod to remove the module:

```
$ ifconfig eth2 down
$ rmmod nx_lsa
```

The installed driver may be removed from the system as follows:

Remove nx_lsa.ko from /lib/modules/<kernel_version>/kernel/drivers/net.

Note: The kernel_version can be obtained by running *uname -r*

To unload the LSA driver, first remove the offload rules. The current list of rules can be displayed by typing **`nxoffload -s`** at the command prompt.

Once all of the offload rules have been removed, the lsa driver can be removed by typing: **`rmmod nx-lsa`**

5.7.4 Removing the intercept driver

The intercept driver cannot be removed with the **`rmmod`** command. Remove the file nx_intercept.ko from /lib/modules/<Kernel Version>/kernel/drivers/net. Then reboot the system.

5.8 Bonding

Refer to section 2.5.4, “Linux bonding” on page 14 for an in-depth discussion of bonding concepts in the IBM BladeCenter.

5.8.1 NIC bonding on Red Hat Linux

The following is based on information found in the *bonding.txt* file located in: /usr/share/doc/kernel-doc-<kernel-version>/Documentation/networking/bonding.txt

Initially, the bonding driver needs to be loaded. This can be accomplished at the command line, or by adding the commands to the /etc/modeprobe.conf file as shown in Example 5-4 where ethX refers to Eth2 or Eth3.

Example 5-4 Changes to /etc/modprobe.conf

```
alias bond0 bonding
options bond0 mode=1 miimon=100 updelay=135000 primary=eth2
```

The updelay parameter places a delay on the failback to allow the switch time to become fully activated after restart. This information is based on RETAIN® Tip # H187467. Based on our testing with the 10 Gb Ethernet Switch Module, values of updelay as low as 20000 ms (20 seconds) will allow the switch sufficient time to reboot and for the network to converge. Using

no updelay appears to work in Layer 3 configurations where VRRP is configured on the switches for the VLAN(s) which connect to the Linux blades. It is always good practice to verify that your environment is properly providing end-to-end High Availability as intended.

Next, the bond0 interface parameters are configured in the ifcfg-bond0 file, as shown in Example 5-5.

Example 5-5 /etc/sysconfig/network-scripts/ifcfg-bond0

```
BOOTPROTO="static"
BROADCAST="10.10.10.255"
IPADDR="10.10.10.107"
NETMASK="255.255.255.0"
NETWORK="10.10.10.0"
REMOTE_IPADDR=""
STARTMODE="onboot"
BONDING_MASTER="yes"
BONDING_SLAVE0="eth2"
BONDING_SLAVE1=”eth3"
```

Note: Replace the sample BROADCAST, IPADDR, NETMASK, and NETWORK values with the appropriate values for your network.

The interfaces that will participate in the bond need to be configured appropriately for the bonding configuration, as shown in Example 5-6 and Example 5-7.

Example 5-6 Changes to /etc/sysconfig/network-scripts/ifcfg-eth2

```
DEVICE=eth2
TYPE=Ethernet
USERCTL=yes
BOOTPROTO=none
MASTER=bond0
SLAVE=yes
```

Example 5-7 Changes to /etc/sysconfig/network-scripts/ifcfg-eth3

```
DEVICE=eth3
TYPE=Ethernet
ONBOOT=yes
USERCTL=yes
BOOTPROTO=none
MASTER=bond0
SLAVE=yes
```

After making the changes as required, restart the networking service. This can be accomplished by typing:

```
$ service network restart
```

5.8.2 Bonding in Red Hat Linux Enterprise 5

Bonding in Red Hat Linux Enterprise version 5 is exactly the same as RHEL 4. Therefore, a detailed example is not included here.

5.8.3 Bonding in SUSE Linux 10

The following information is based on Novell's TID# 3929220 located at:
https://secure-support.novell.com/KanisaPlatform/Publishing/133/3929220_f.SAL_Public.html

Initial setup of the NetXen 10 Gb card

Novell recommends initially configuring the Ethernet adapter using the Yast utility. This can be accomplished from the command line or from the graphical interface. Set up the Netxen interfaces to use DHCP (the default option). Once complete, there should be two more Ethernet interfaces available on the server; Eth2 and Eth3. Eth0 and Eth1 are the on-board broadcom interfaces, which can be seen by typing **ifconfig** at the console.

SUSE Linux Enterprise Server (SLES) uses a slightly different naming convention for the Ethernet interface configuration files. Instead of using ifcfg-eth2, as Red Hat does, a file is used called ifcfg-eth-id-xx:xx:xx:xx:xx where xx represents the hexidecimal MAC address of the Ethernet card. It is also possible to have a filename that uses the PCI bus ID for the file name, as in ifcfg-eth-bus-pci-xxxx.xx:xx.x.

Additionally, SLES identifies the PCI bus interface for each Ethernet adapter. Once Yast has completed the basic configuration of the network adapters, open the file for editing. The file should look something like Example 5-8.

Example 5-8 /etc/sysconfig/network/ifcfg-bond0

```
BOOTPROTO='static'
BROADCAST='10.10.10.255'
IPADDR='10.10.10.101'
NETMASK='255.255.255.0'
NETWORK='10.10.10.0'
STARTMODE='onboot'
BONDING_MASTER='yes'
BONDING_SLAVE0='bus-pci-0000:0c:00.2'
BONDING_SLAVE1='bus-pci-0000:0c:00.3'
```

It is not possible to assign a primary interface to the bond when creating the bonded interface in the ifcfg-bond0 config file. Creating the bond interface in the /etc/modprobe.conf.local file will accomplish this task (see Example 5-9).

Example 5-9 Additions to /etc/modprobe.conf.local on SLES10 to create a bond and assign a primary interface

```
alias bond0 bonding
options bond0 mode=1 miimon=100 primary=eth3
```

Example 5-10 and Example 5-11 on page 95 show the contents of each Ethernet ID.

Example 5-10 Contents of ifcfg-eth-id-00:0e:1e:00:12:2a

```
BOOTPROTO='none'
STARTMODE='off'
BROADCAST=''
ETHTOOL_OPTIONS=''
IPADDR=''
MTU=''
NAME='Ethernet controller'
NETMASK=''
```

```
NETWORK=''
REMOTE_IPADDR=''
STARTMODE='auto'
UNIQUE='gvUW.Jtk7SnUeMY4'
USERCONTROL='no'
_nm_name='bus-pci-0000:0c:00.2'
```

Example 5-11 Contents of ifcfg-eth-id-00:0e:1e:00:12:2b

```
BOOTPROTO='none'
BROADCAST=''
ETHTOOL_OPTIONS=''
IPADDR=''
MTU=''
NAME='Ethernet controller'
NETMASK=''
NETWORK=''
REMOTE_IPADDR=''
STARTMODE='auto'
UNIQUE='ZaZ6.Jtk7SnUeMY4'
USERCONTROL='no'
_nm_name='bus-pci-0000:0c:00.3'
```

6

Nortel 10 Gb Ethernet Switch Module configuration and network integration

This chapter outlines the switching topologies used in the writing of this Redpaper.

6.1 Summary of sample configurations included

The sections in this chapter present several sample configurations which exploit different capabilities of the Nortel 10 Gb Ethernet Switch Module for IBM BladeCenter. It is not intended that any of these samples be copied exactly and used in a real network. However, portions of one or more of them can serve as the basis for a real configuration. In general, the configurations shown were tested on the Nortel 10 Gb switch module, except where noted. However, the configurations can be readily and easily adapted for use on other Nortel switch modules including those which run solely at 1 Gb. Similarly, the configurations in the Redpaper, *Nortel Networks L2/3 Ethernet Switch Module for IBM eServer BladeCenter*, REDP-3586, can be adapted for use on the Nortel 10 Gb ESM and Nortel 10 Gb Uplink ESM, although the syntax of some commands has changed since the earlier Redpaper was published.

6.1.1 Layer 2 configuration summary

The Layer 2 configurations shown below all rely on one of the versions of Spanning Tree Protocol (STP). To provide redundancy, the Layer 2 topologies shown have multiple paths available between switches. This creates Layer 2 loops, which will quickly create broadcast storms which will consume much of the available bandwidth and processing power on the switches involved. All of the variants of STP function by blocking some ports - preventing traffic from being send on them - to break these Layer 2 loops. The newer STP standards differ from the original in the speed of convergence after a topology change and the ability to support multiple VLANs.

The topologies shown include the use of static and LACP trunks (called channels, Etherchannel, or Port Channels by Cisco) between Nortel and Cisco switches. Single links as well as trunks are shown using 802.1q VLAN tagging (Cisco “trunk” ports.)

One topology is specifically for the deployment of multiple 10 Gb Uplink switches in multiple chassis and shows how to interconnect these switches and chassis with one 10 Gb link to the core network shared across several chassis. This topology is shown in 6.5, “Stacking configuration with 10 Gb Uplink switches” on page 129.

The configurations shown provides sufficient redundancy to protect against some single point failures. All of the configurations work with NIC teaming (called “bonding” in Linux) to enable a more robust high availability solution. A discussion of teaming and bonding can be referenced in 2.5, “NIC bonding” on page 13.

In our testing of Layer 2 configurations we again encountered interoperability issues between the 10 Gb Nortel switch module Cisco IOS running on Cisco 65xx switches. We encountered similar issues, also involving Multiple Spanning Tree, during the writing of the *Nortel Networks L2/3 Ethernet Switch Module for IBM eServer BladeCenter*, REDP-3586. The version of IOS we used on the 65xx switches was compiled in late 2006; however, there are other versions which do provide standards-compliant support for the MST protocol.

6.1.2 Layer 3 configuration summary

The Layer 3 topologies shown all implement redundancy and high availability without blocking any of the ports or links between switches. Broadcast storms of the kind that can cause problems in Layer 2 configurations almost never occur in a Layer 3 configuration because broadcast traffic is contained within a subnet/VLAN with some exceptions which must be explicitly configured.

In order to do this, routing between subnets/VLANs is used and must be configured. Each switch must have a routing table showing where to forward traffic based on the destination's IP address. These tables can be configured manually (static routing) or learned dynamically from other routing switches (dynamic routing). Both options are shown; two different dynamic routing protocols (RIP and OSPF) are shown.

High availability is achieved in these designs through the use of Virtual Router Redundancy Protocol (VRRP), Trunk Failover, and NIC teaming/bonding. VRRP allows the two Nortel 10Gb ESM modules to back each other up, so that the failure of one Nortel 10Gb ESM can be survived.

The decision to use static or dynamic routing is made by network architects in most cases. These configurations are provided to show that the Nortel 10Gb ESM switch modules can participate in RIP or OSPF networks if it is decided that this is the best way to integrate them with the existing network.

6.2 Basic configuration, options, and issues

Before discussing the specifics of each configuration, it is necessary to outline the hardware and software used during the experiments.

6.2.1 Hardware and software used for the lab environment

It should be noted that the choice of the elements of the tested configurations was made based on the assumption that the BladeCenter is being deployed in a mission-critical data center environment, where high availability and performance are of utmost importance. Also, the Cisco 6500 switches support the standards that are used in each experiment for Layer 2 and 3 switching.

BladeCenter H chassis configuration

The IBM BladeCenter H chassis was configured as follows:

One BladeCenter H chassis (8852) with:

- Two Nortel 10 Gb Ethernet Switch Module for IBM BladeCenter (39Y9267)
- One HS21 (8853) Intel blade server
  - One Netxen 10 Gb Ethernet CFFh card
  - Microsoft Windows Server 2003 32-bit Standard Edition with Service Pack 2
- One HS21 (8853) Intel bladeserver
  - One Netxen 10 Gb Ethernet CFFh card
  - Red Hat Enterprise Linux version 4 update 4
- One LS41 (7972) AMD™ bladeserver
  - One NetXen 10 Gb Ethernet CFFh card
  - Novell SUSE Linux Enterprise Server SLES version 10
- One LS41 (7972) AMD bladeserver
  - One NetXen 10 Gb Ethernet CFFh card
  - Microsoft Windows Server 2003 R2 64-bit Enterprise

BNT/Nortel switches

- 3 Nortel L2/3 10 Gb Uplink Switch Modules running version 1.1.0.68

- 2 Nortel 10 Gb Switch Modules running version 1.0.1.1

Cisco upstream switches

- 1 Cisco 6509 running IOS version 12.2(18)SXD7b
- 1 Cisco 6506 running IOS version 12.2(18)SXD7b

6.2.2 Preconfiguration preparation

The configurations in this document were built off each other in a progressing manner. Each example is a progression from the first. Configurations after each experiment were not reset to start from scratch.

Important: If working in a production network, be sure to understand the consequences of any commands that are issued. Failure to completely understand the operation of commands can lead to network down conditions.

Note: Available features and command syntax can be different with different versions of code. This document was prepared using the features and syntax from the aforementioned revisions of code, and as such, might vary from other revisions. For complete and current lists of available features and commands for these products, visit the IBM or Nortel Web sites.

6.2.3 Configuration parameters common to all examples

This section lists some established configuration options that are common to all of the examples. These are only for demonstration purposes in the examples and might or might not be duplicated in your particular environment.

Management Module settings for Nortel 10 Gb Ethernet Switch Modules

Each Nortel 10 Gb Ethernet Switch Module is configured with an IP address for the MGT1 ports:

- Nortel 10Gb ESM 1 is located in bay 7 and is configured with 9.42.170.246 and mask 255.255.254.0
- Nortel 10Gb ESM 2 is located in bay 9 and is configured with 9.42.170.247 and mask 255.255.254.0
- The default gateway is set to the internal interface of the Management Module 9.42.171.73
- Both Nortel 10Gb ESMs have Fast POST enabled
- Both Nortel 10Gb ESMs have External Ports enabled
- Both Nortel 10Gb ESMs have External Management over all Ports enabled
- Both Nortel 10Gb ESMs have Preserve new IP configuration on all resets enabled

We do not discuss the Management Module configuration from this point forward. There are no changes to the settings listed above.

IP address and VLAN scheme

NetXen 10 Gb Ethernet interfaces were used on VLAN 10 (and sometimes VLAN 3) with IP addresses 10.10.10.x (and 10.10.3.x) and subnet mask 255.255.255.0. The onboard Broadcom 1 Gb interfaces received addresses through DHCP and were used for remote

access to the servers for management and reconfiguration (IP address 9.42.170.x and mask 255.255.254.0).

As mentioned in the previous section, switch management was done using the 9.42.170.x subnet. The Cisco 6509 was set with address 9.42.170.243. Due to a lack of static IP addresses to assign on this subnet, the Cisco 6506 was managed through an interface connected to the 6509 using VLAN 3 (10.10.3.201).

Other IP address modifications are covered as the examples evolve in the sections that follow.

6.3 Layer 2 configurations

This section covers topologies using best practiced layer 2 standards. These same standards are covered in greater detail in *Nortel Networks L2/3 Ethernet Switch Module for IBM eServer BladeCenter*, REDP-3586.

6.3.1 Layer 2 standards and technologies covered

This section provides a brief overview of the Layer 2 standards covered in the experiments to follow.

VLAN tagging - 802.1Q

This standard defines the use of a tag field in the header of each packet which identifies the VLAN to which the packet belongs. This feature on the Nortel 10 Gb Switch Module is configured with the **`/cfg/port/tag ena|dis`** command.

Cisco refers to this as a *trunk* and uses **`switchport trunk`access** command to configure it.

Link aggregation - LACP - 802.3ad and 802.3-2002

These standards define techniques for grouping two or more parallel connections between a pair of devices as a single logical link with the total bandwidth of its members. Static trunk is configured with the **`/cfg/l2/trunk`** command. LACP is configured with the **`/cfg/l2/lacp`** command.

Cisco refers to this as a *Port Channel* or *EtherChannel* and uses the **`channel-group`** and **`interface portchannel`** commands to configure it.

Spanning tree - 802.1D, 802.1w, 802.1s

These protocols define techniques for managing Layer 2 networks whose topologies include loops. They ensure that a broadcast packet (or any other packet) is not forwarded endlessly around such a loop by logically blocking some ports.

Classic STP is defined by the 802.1D standard. This is the original Spanning Tree standard. The Nortel will interoperate with Cisco using Cisco's proprietary extension to 802.1D, known as Per-VLAN Spanning Tree (PVST+).

Two of the key shortcomings of classic STP is that it takes as much as 50 seconds to recover from the failure of a link or device; and that it does not deal well with multiple VLANs carried over the same physical link (802.1Q). These are remedied by 802.1w standard for Rapid Spanning Tree (RST or RSTP) and by the 802.1s standard for Multiple Spanning Tree (MST or MSTP).

The Spanning Tree functions on the Nortel 10 Gb Switch Module are configured from the /cfg/l2/stg and /cfg/l2/mrst menus.

It is possible to disable all Spanning Tree functionality between the Nortel switch modules in the BladeCenter chassis and the upstream Cisco (in these tests) or other switches. This is because the Nortel switches in the BladeCenter do not pass traffic to each other within the chassis except over the isolated management module VLAN (4095). As long as the switches are not cabled to each other there is no possibility of a loop. In order to do this, one can disable Spanning Tree on the ports on the Nortel and Cisco switches that connect them to each other. On the Nortel switches it is possible to turn Spanning Tree off entirely. This should only be done if the entire topology of the network is well understood and the personnel involved are sure that it is safe.

Trunk failover for high availability

Trunk failover is covered in 2.5.5, “Trunk failover for high availability” on page 16.

Trunk failover is configured on the Nortel 10 Gb Switch Module according to Example 6-1.

Example 6-1 Enabling trunk failover

```
/cfg/l2/failover/on
   trigger <n>
   ena
   amon
      addtrnk <static trunk group number - once for each trunk to be monitored>
      -OR-
      addkey <LACP adminkey value - once for each LACP key to be monitored>
```

If there are multiple trunk groups which are critical upstream connections, such as to multiple upstream switches, then they should all be included in one of the triggers. It is possible to configure the *limit* option on a trigger which sets the minimum number of remaining ports that will trigger failover. (The default is 0 which means all ports configured in the trigger must be down; a value of 1 would mean that 2 surviving ports will not cause failover to be triggered but only 1 port *will* fail; presumably only one port would not be sufficient to meet the bandwidth required). See the TechTip at:
http://www.redbooks.ibm.com/abstracts/tips0597?Open
for details on how this is done.

In most cases, you should configure trunk failover on all Nortel 10 Gb Switch Modules in the BladeCenter if the server blades are running NIC Teaming.

6.3.2 Layer 2 configuration with 802.1D, Cisco PVST+ and Nortel extensions

The topology in this test was an “inverted U” topology, where each Nortel switch is connected to a corresponding Cisco 6500. VLAN tagging (802.1Q), static trunks, and LACP were used between the Nortel and Cisco switches.

The connection between 10GESM1 and C6509 is a two-port static trunk. The connection between 10GESM2 and C6506 is a two-port LACP configured trunk. VLAN tagging (802.1Q) is enabled across both of the trunks; VLANs 3 and 10 are carried across all of the switches.

This test shows interoperability between Nortel and Cisco switches using classic 802.1D Spanning Tree and proprietary extensions from both Nortel and Cisco. These extensions are necessary to enable 802.1D to support environments where multiple VLANs are in use and to improve upon the slow convergence of 802.1D.

On the Cisco switches, Per Vlan Spanning Tree (PVST+) was used; this feature sends Spanning Tree information (Bridge Protocol Data Units, BPDUs) on each and every VLAN on a link which carries multiple VLANs. The corresponding feature on the Nortel side is multiple Spanning Tree Groups, which will send BPDUs on any desired VLANs where it is configured. These features interoperate successfully.

In addition, on the Nortel side the *fast port* feature was enabled on the external ports which were connected to the Cisco switches. This allows the ports on the Nortel switches to go directly to Spanning Tree *forwarding (FWD)* state rather than progressing through the *listening* and *learning* states and taking 50 seconds to reach *forwarding* state.

A diagram of the topology of this test is shown in Figure 6-2, and the full configuration is shown in A.1.2, “Configurations for 6.3.2, “Layer 2 configuration with 802.1D, Cisco PVST+ and Nortel extensions” on page 102” on page 148.

This topology is commonly used and provides protection against failures of:

- One of the Nortel switches
- The connection from one switch to the blade server, across the chassis midplane (very rare)
- One of the Cisco core switches
- Both links of the connection between one Nortel switch module and the corresponding Cisco switch
- One link of either or both connections between each Nortel switch and the corresponding Cisco switch

The only single point of failure in this network topology is the cross-link between the two Cisco switches, and that could easily be remedied in a customer environment. It did not have any impact on our tests.

The only drawback of this configuration is that it is possible that some of the ports on the Cisco switches will still go through the 50 second delay in the event of a topology change (typically caused by a link or device failure). The ports on the Cisco switches should not be set to go to *forwarding* state directly because the Cisco switches are not at the edge of the network.

The configuration commands for the Cisco switches to enable PVST+ and for the Nortel switches to enable multiple groups and use *fast forwarding* are shown in Example 6-2 and Example 6-3.

Example 6-2 Cisco commands to enable PVST+

```
spanning-tree mode pvst
```

Example 6-3 Nortel commands to enable multiple STP groups and Fast Forward

```
/c/l2/stg 1/on /* stg 1 gets all of the vlans not explicitly placed elsewhere */
/c/l2/stg 2/on
   add <vlan>
/c/l2/stg 3/on
   add <vlan>
...etc...
/c/port <extn> /* for each external port where this is desired */
   fastfwd ena
```

For more instruction on how to build the topology without creating loops, as well as the command set for enabling each standard (on both IOS and AOS) refer to *Nortel Networks L2/3 Ethernet Switch Module for IBM eServer BladeCenter*, REDP-3586.

6.3.3 Layer 2 configuration with MSTP

In order to provide uniformly rapid convergence after a topology change, the configuration above was modified to use Multiple Spanning Tree (MST, MSTP, 802.1s) rather than 802.1D.

The topology is shown in Figure 6-1.

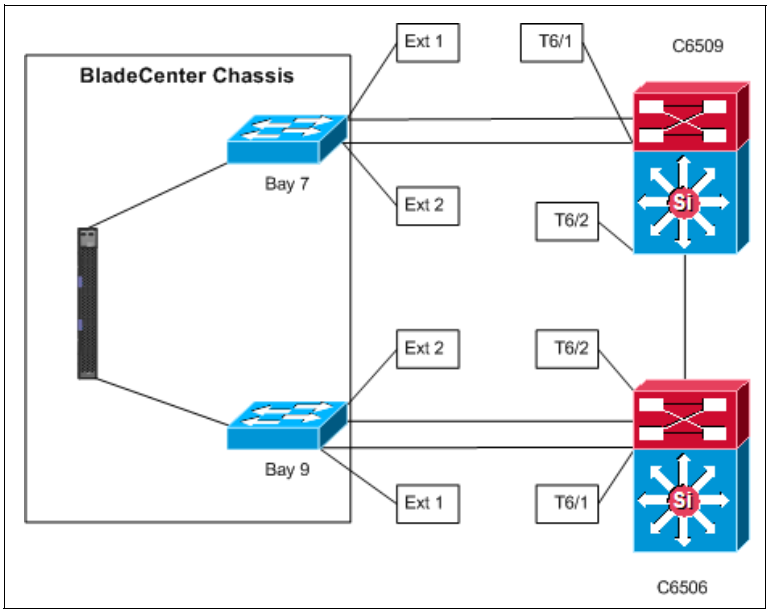


Figure 6-1 L2 configuration with VLAN tagging, static trunk, LACP, and MRST

Our initial intent was to use MSTP on both Cisco and Nortel switches, using multiple instances of MSTP even though there were no differences in the topologies of the VLANs in use to verify that the protocol works and interoperates successfully. What we observed was that despite several attempts to set the configuration parameters on the switches, the Cisco switches always saw the ports connecting to the Nortel switches as *boundary* ports to a region where the RSTP protocol is running. This is normal behavior for MSTP when there is a MSTP parameter mismatch, which can be done deliberately to partition a network into multiple MSTP regions. In our case there was no mismatch. We therefore decided to capture the MSTP traffic (BPDUs) to better understand what was happening.

Port mirroring was set up on the Nortel side according to Example 6-4.

Example 6-4 10GESM1: Set up port mirroring for EXT1 and EXT2

```
/c/pmirr/monport INT6
        add EXT1 both
/c/pmirr/monport INT6
        add EXT2 both
```

Ethereal was used on a Windows 2003 SE installed blade with a 10 Gb NIC and a trace was done on this topology. During the trace it was observed that the decoded BPDUs from the Cisco 6500s were labeled *malformed*. The impact of this is that the Nortel switches would

show as boundary switches, meaning they were neighbors, but not in the same MSTP region. This makes for an unstable topology.

Note: The Tolly Group is an independent third-party testing organization currently doing a study of interoperability between Nortel and Cisco switches. The Tolly Group was able to successfully build this configuration using an IOS version that was not used in this paper. The Cisco IOS version that was verified by The Tolly Group was: s72033-ipservicesk9-mz.122-18.sxf4. The full interoperability report from Tolly is available at: http://www.tolly.com/DocDetail.aspx?DocNumber=206168

This configuration would solve the convergence issue identified previously if we were able to test with the IOS version mentioned in the note above. When implementing MSTP, be aware of ports identified as *boundary RSTP* ports and verify that any partitioning of the network into multiple MSTP regions is intentional.

The commands to configure MSTP on the Cisco switches is shown in Example 6-5.

Example 6-5 Cisco commands to configure MSTP

```
spanning-tree mode mst
spanning-tree mst config
name TENGIG
revision 1
instance 3 vlan 3
instance 2 vlan 10
...etc....if needed for other VLANs
/* instance 0 is automatically configured with all other vlans 1-4094 which are
not explicitly placed elsewhere */
```

Example 6-6 shows the commands to configure MSTP on the Nortel switches.

Example 6-6 Nortel commands to configure MSTP

```
/c/l2/mrst/on
   mode mstp
   rev 1
   name "TENGIG"
   cist/add 1 /* this is the equivalent to Cisco instance 0 */
      port EXT1/cost 1000 /* force 1Gb link between Cisco switches to BLK first*/
      port EXT2/cost 1000 /* this is optional */
/c/l2/stg 2/clear /* equivalent to instance 2 */
   add 10
   port EXT1/cost 1000
   port EXT2/cost 1000
/c/l2/stg 3/clear
   add 3
   port EXT1/cost 1000
   port EXT2/cost 1000
```

Full configuration dumps are located in A.1.1, “Configurations for 6.3.3, “Layer 2 configuration with MSTP” on page 104” on page 138.

6.3.4 Layer 2 configuration with MSTP and Nortel Multiple STP groups

This test was conducted to overcome the previous findings where decoded BPDUs were being shown as malformed in Ethereal.

Figure 6-2 shows a near-full mesh topology, which is different from the topologies used in the previous examples. There are no more aggregated links between the switches; instead the crisscross links shown are used, connecting each Nortel switch to both of the upstream Cisco 6500s.

Spanning Tree blocked the crisscross links, as expected, to prevent looping. This is common practice, but it must be understood that the blocked links will not carry any production traffic unless there is a failure to a live link. If a failure to a live link occurs, MSTP will recover in 2 seconds. For many network solutions this 2 second downtime period of recovery is perfectly acceptable.

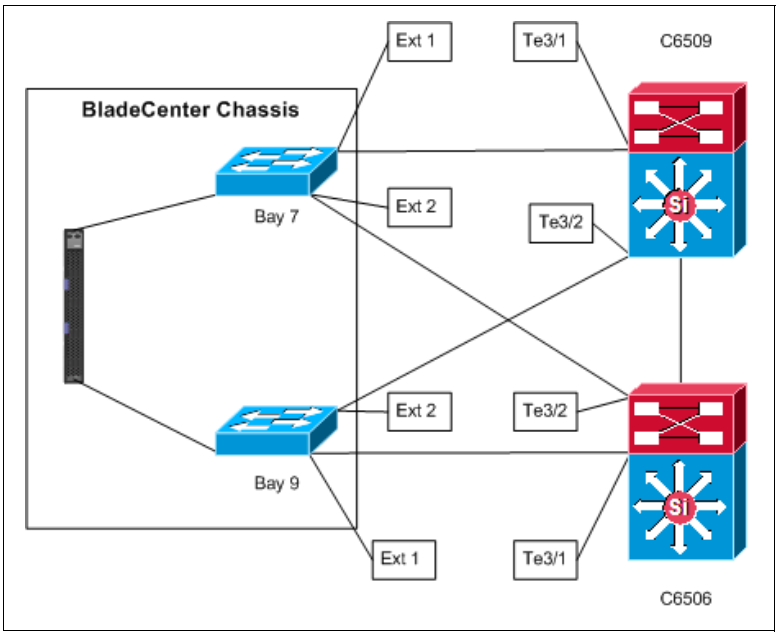


Figure 6-2 L2 configuration with MST on Cisco switches and Fast Forwarding on Nortel switches

Multiple Spanning Tree groups (which function similarly to Cisco PVST+ and interoperate with it) are used on the Nortel switches. Since MSTP is meant to work with classic and rapid Spanning Tree, we chose to use classic 802.1D on both Nortel switches.The MSTP functionality on the Cisco switches properly identified the ports connecting to the Nortel switches as *boundary PVST+* ports. The ports on the Nortel switches which connect to the Cisco switches are configured with *fast port* as was done in the first example which used only 802.1D, to provide rapid convergence on the Nortel side. MSTP provides rapid convergence on the Cisco side.

The configuration commands for STP and MSTP are shown in “Spanning tree - 802.1D, 802.1w, 802.1s” on page 101 and 6.3.2, “Layer 2 configuration with 802.1D, Cisco PVST+ and Nortel extensions” on page 102.

6.3.5 Bonding and Teaming in a Layer 2 topology

This Spanning Tree configuration was tested with the bonding driver of RHEL4 Update 4. This driver provides equivalent function to the NetXen and Broadcom NIC teaming drivers used in Windows environments.

In order to inject a failure, 10GESM1 was powered off through the Management Module. Example 6-7 shows a watch that was conducted on bond0. It can be seen that eth3 becomes active after the 10GESM1 was powered off.

Note that this test was performed *without* the use of the *fast port* feature on the Nortel switches, as can be seen in the PING test below.

Example 6-7 Watch conducted on bond0

```
Ethernet Channel Bonding Driver: v2.6.3 (June 8, 2005)
Bonding Mode: fault-tolerance (active-backup)
Primary Slave: eth2
Currently Active Slave: eth2
MII Status: up
MII Polling Interval (ms): 100
Up Delay (ms): 135000
Down Delay (ms): 0
Slave Interface: eth2
MII Status: up
Link Failure Count: 0
Permanent HW addr: 00:0e:1e:00:13:c0
Slave Interface: eth3
MII Status: down
Link Failure Count: 0
Permanent HW addr: 00:0e:1e:00:13:c1

[root@hs21-rhel4 network-scripts]# cat /proc/net/bonding/bond0
Ethernet Channel Bonding Driver: v2.6.3 (June 8, 2005)
Bonding Mode: fault-tolerance (active-backup)
Primary Slave: eth2
Currently Active Slave: eth3
MII Status: up
MII Polling Interval (ms): 100
Up Delay (ms): 135000
Down Delay (ms): 0
Slave Interface: eth2
MII Status: down
Link Failure Count: 1
Permanent HW addr: 00:0e:1e:00:13:c0
Slave Interface: eth3
MII Status: up
Link Failure Count: 0
Permanent HW addr: 00:0e:1e:00:13:c1

[root@hs21-rhel4 network-scripts]# cat /proc/net/bonding/bond0
Ethernet Channel Bonding Driver: v2.6.3 (June 8, 2005)
Bonding Mode: fault-tolerance (active-backup)
Primary Slave: eth2
Currently Active Slave: eth2
MII Status: up
```

```
MII Polling Interval (ms): 100
Up Delay (ms): 135000
Down Delay (ms): 0
Slave Interface: eth2
MII Status: up
Link Failure Count: 1
Permanent HW addr: 00:0e:1e:00:13:c0
Slave Interface: eth3
MII Status: up
Link Failure Count: 0
Permanent HW addr: 00:0e:1e:00:13:c1
[root@hs21-rhel4 network-scripts]#
```

Example 6-8 shows the output of dmesg on bond0. Some additional text was added showing when 10GESM1 was powered off and then back on. From the output it can be seen how the bonding driver is handling the failover and failback.

Example 6-8 dmesg on bond0

```
Ethernet Channel Bonding Driver: v2.6.3 (June 8, 2005)
bonding: MII link monitoring set to 100 ms
divert: allocating divert_blk for bond0
ip_tables: (C) 2000-2002 Netfilter core team
netxen_nic eth2 (port 0), Link is up
bonding: bond0: making interface eth2 the new active one 0 ms earlier.
bonding: bond0: enslaving eth2 as an active interface with an up link.
netxen_nic eth3 (port 1), Link is up
bonding: bond0: enslaving eth3 as a backup interface with an up link.
bonding: bond0: link status definitely up for interface eth3.

/**** 10GESM1 powered off at this point ****/

netxen_nic: eth2 NIC Link is down
bonding: bond0: link status definitely down for interface eth2, disabling it
bonding: bond0: making interface eth3 the new active one.

/**** 10GESM1 powered on at this point ****/

netxen_nic: eth2 NIC Link is up
bonding: bond0: link status up for interface eth2, enabling it in 135000 ms.
bonding: bond0: link status definitely up for interface eth2.
bonding: bond0: making interface eth2 the new active one.
```

Example 6-9 shows a ping that was done during this experiment. The gap between sequence number 33 and 65 is the time that Spanning Tree took to recover. There is a second gap between sequence number 166 and 196 where failback occurred.

Example 6-9 Pings

```
[root@hs21-rhel4 ~]# ping 10.10.10.201
PING 10.10.10.201 (10.10.10.201) 56(84) bytes of data.
64 bytes from 10.10.10.201: icmp_seq=0 ttl=255 time=0.513 ms
64 bytes from 10.10.10.201: icmp_seq=1 ttl=255 time=0.319 ms
64 bytes from 10.10.10.201: icmp_seq=2 ttl=255 time=0.284 ms
64 bytes from 10.10.10.201: icmp_seq=3 ttl=255 time=0.266 ms
......(no missing pings)
```

```
64 bytes from 10.10.10.201: icmp_seq=32 ttl=255 time=0.290 ms
64 bytes from 10.10.10.201: icmp_seq=33 ttl=255 time=0.295 ms
64 bytes from 10.10.10.201: icmp_seq=65 ttl=255 time=0.341 ms
64 bytes from 10.10.10.201: icmp_seq=66 ttl=255 time=0.339 ms
......(no missing pings)
64 bytes from 10.10.10.201: icmp_seq=165 ttl=255 time=0.310 ms
64 bytes from 10.10.10.201: icmp_seq=166 ttl=255 time=0.309 ms
64 bytes from 10.10.10.201: icmp_seq=196 ttl=255 time=0.348 ms
64 bytes from 10.10.10.201: icmp_seq=197 ttl=255 time=0.385 ms
64 bytes from 10.10.10.201: icmp_seq=198 ttl=255 time=0.311 ms
```

To enhance the speed of recovery to match the MST configured Cisco switches, *Fast Forwarding* was enabled on the Nortel switches. Example 6-10 shows how to enable Fast Forwarding on the EXT ports of the Nortel switches.

Example 6-10 Fast Forwarding enabled on both Nortel switches

```
/c/port EXT1
fastfwd ena
/c/port EXT2
fastfwd ena
```

After setting Fast Forwarding on the Nortel switches the experiment was done over and no icmp_seq gaps were seen during the ping.

Note: If MSTP is the choice solution, it is recommended to use the Cisco IOS version s72033-ipservicesk9-mz.122-18.sxf4. If that is not possible, it is recommended to use the solution presented in this section that uses classic Spanning Tree (802.1D) with Fast Forwarding on the Nortel switches with MST enabled on the upstream Cisco switches.

Full configuration dumps are located in A.1.2, “Configurations for 6.3.2, “Layer 2 configuration with 802.1D, Cisco PVST+ and Nortel extensions” on page 102” on page 148.

6.4 Layer 3 configurations

This section covers topologies using best practice Layer 3 standards. These same standards are covered in greater detail in *Nortel Networks L2/3 Ethernet Switch Module for IBM eServer BladeCenter*, REDP-3586.

For more instruction on how to build the topology without creating loops, as well as the command set for enabling each standard (on both IOS and AOS) refer to *Nortel Networks L2/3 Ethernet Switch Module for IBM eServer BladeCenter*, REDP-3586.

6.4.1 Layer 3 topologies

Our testing included two basic topologies at Layer 3, just as it did at Layer 2. These two topologies are recommended but are not the only ones possible. They both include two Nortel switch modules in each BladeCenter chassis (although our testing was done on only one chassis) connected to two upstream switches. Depending on the robustness of high availability desired, it is possible to connect two switch modules in a BladeCenter chassis to a single upstream switch, or even to use only one switch module in a chassis. (If there is only a

single upstream switch used it is recommended that ports on different line cards be used to connect to the switch module(s) in the chassis).

The two topologies used are labeled as follows:

- The “Inverted U” where each switch module is connected to a corresponding upstream switch with one or more links in a trunk
- Full (or nearly full) mesh, where each switch module is connected to the two upstream switches with one or more links in a trunk

In both the Layer 3 topologies, the two upstream (Cisco) switches are connected to each other and the two Nortel switch modules are cross-connected to each other.

It is also possible to implement a partial Layer 3 topology which does not require the Nortel switch modules to be cross-connected.

6.4.2 Modifications to move from L2 to L3

Layer 3 configurations in this section are different from the Layer 2 configurations in the following ways:

- 802.1q trunking is not required on the connections to the Cisco switches. Traffic from the VLANs used for the server blades is sent on the uplink ports as routed traffic and identified by its source and destination IP addresses. There is no need to identify its VLAN association. The ports on the Cisco switches are therefore configured as *switchport access* or as *routed* ports. The ports on the Nortel switches are configured as *untagged* ports which is equivalent to Cisco *access* ports.
- Additional VLANs and associated subnets are required for the connections to the Cisco switches. The uplink connections do not share the VLANs or subnets. They are also different from the subnets used by the blade servers. The VLANs and subnets used are:
  - VLAN 5: 10.10.5.x, used to connect 10GESM1 to C6509
  - VLAN 15: 10.10.15.x, used to connect 10GESM1 to C6506
  - VLAN 20: 10.10.20.x, used to connect 10GESM2 to C6506
  - VLAN 25: 10.10.25.x, used to connect 10GESM2 to C6509
  - VLAN 30: 10.10.30.x used to connect C6509 to C6506
- Upstream routers must be able to find the route to the blade servers. This requires either the use of explicitly configured (static) routes or the use of a dynamic routing protocol. Dynamic routing protocols enable routers to share route information with their neighbors. Examples of the use of static and dynamic routing are included in this section.
- High availability designs using Layer 3 require different techniques than those using solely Layer 2. Examples of HA designs are included in this section.

Note: As mentioned previously, *Nortel Networks L2/3 Ethernet Switch Module for IBM eServer BladeCenter*, REDP-3586 includes more information on the specific L2 and L3 standards used.

6.4.3 Reasons for choosing a L3 topology over a L2 topology

The reasons for choosing a Layer 3 design for blade server switching are outlined below.

- Layer 3 switching keeps more traffic within the BladeCenter chassis.

  In any design where the blade servers are assigned to multiple VLANs, routing is required for servers which are not on the same VLAN to communicate with each other. (An

example of such a design would be using blade servers as WebSphere® Web and application servers and placing the Web servers on one VLAN and the application servers on a different VLAN.)

With a Layer 3 configuration on the 10GESM, servers on different VLANs can communicate through the switch module which is inside the BladeCenter chassis. If only Layer 2 switching is used within the chassis, then traffic between, for example, WebSphere Web and application servers would leave the chassis on the external links and flow through one or more external devices until it reached a Layer 3 switch (router). It would then flow through one or more additional devices until it returned to the BladeCenter chassis, crossing the external links for a second time.

The benefits of keeping traffic within the chassis include greater security - because there are no patch panels which can be tampered with - and lower latency, since the traffic traverses a smaller number of switches between source and destination.

- Layer 3 switching allows more efficient use of external connections

  The key issue here is that use of Layer 3 allows the network to run without the use of Spanning Tree Protocol (STP). Spanning tree works by blocking links which would create a loop; connections from a 10GESM to two or more upstream switches which are connected to each other fall into this category. The consequence is that up to half of the links from the 10GESM would be *blocked* during normal operations and would not carry traffic. Only if the active link(s) failed would the blocked links be used to carry traffic. Upon those link failures, depending on the Spanning Tree protocol used, the wait-time for recovery may not be acceptable. This is exactly what was seen in the preceding section.

  Layer 3 routing not only allows all of the uplinks to be active but also allows the 10GESM to send traffic to a given destination on the best path to that destination.

6.4.4 Dynamic routing standards

In the Layer 3 examples to follow, both RIP and OSPF dynamic routing protocols are covered.

RIP - RFC1058 and RFC2453

RIP is used by Layer 3 routers to exchange routing table information about the networks which they can reach and determine how far away those networks are. This facilitates end-to-end IP connections which traverse multiple routers.

RFC1058 defines the original RIP specification. This was enhanced with the definition of RIP version 2 in RFC 2453.

RIP is configured on the Nortel 10 Gb Switch Module using the **`/cfg/l3/rip`** menu.

OSPF - RFC1257, RFC2328, and others

OSPF, like RIP, is used by Layer 3 routers to exchange routing table information. It is more scalable and versatile than RIP and recovers from failures more quickly. However, OSPF is more complex and more difficult to configure.

OSPF uses a fundamentally different approach to managing routing tables than RIP in that each router running OSPF maintains a complete representation of the network topology. With RIP, routers are only aware of their immediate neighbors. As a result of this, OSPF requires more memory and more processing power than RIP.

OSPF is configured on the Nortel 10 Gb Switch Module using the **`/cfg/l3/ospf`** menu.

VRRP - RFC 3768

Virtual Router Redundancy Protocol (VRRP) is a Layer 3 protocol used to enable switches to back each other up in a way which is transparent to client and server computers. VRRP works by defining an address which is shared between the switches. One switch which is the Master is the only one which will answer to the shared address. One or more other switches in Backup state are configured to take over from the master in the event of a failure. An instance of VRRP is configured for each VLAN where a shared address is to be used. This implies that if there is one VLAN for the internal ports and an additional VLAN for the external ports, then there can be two instances of VRRP, providing a shared address on the internal VLAN and a different shared address on the external VLAN.

Each switch in a group running VRRP has a configured priority. When VRRP first becomes active, the switch with the highest priority will become the master switch. The master switch sends out periodic hello packets announcing that it is still operational. The backup switch with the highest configured priority will take over when the hello packets are no longer received.

6.4.5 High availability considerations

For high availability (HA) to be truly effective it needs to be well thought out. A complete high availability design should encompass servers, storage, and more of the network than just the portions connected to the BladeCenter chassis. The object is to ensure that there is no single point of failure which can cause the application(s) to become unavailable or unreachable. The topologies in this layer 3 section attempt to show a HA environment. These configurations will survive the following outages:

- Failure of one of the two Nortel 10 Gb switch modules
- Failure of one of the upstream switches
- Failure of the links connecting one Nortel switch to the upstream switch or switches
- Failure of a connection from a server blade to one of the Nortel switch modules (very rare)

The following are some important design considerations to try to ensure connectivity is maintained under various failure scenarios:

- The two Nortel switches should be configured with VRRP to enable Layer 3 high availability. VRRP requires a Layer 2 connection between switches for every VLAN where a VRRP instance is configured. This connection must carry all the VLANs which have a VRRP instance configured. It can be a direct connection or through one or more upstream switches. This connection cannot be through the Management Module VLAN (4095) or the EXT7 (external management port) VLAN (4094).

  Note: Using BladeCenter HT this Layer 2 connection does not need to use any of the EXT ports of the 10GESMs. The connection can be made with ISL interposers.

- The blade servers need to be using the VRRP address(es) for the VLANs where they are configured as their default gateway.
- The blade servers should use NIC teaming/bonding. It is necessary to be sure that the primary NIC in an active-standby team is the one which points to the default VRRP master switch (the switch which is configured with higher priority). Reference 2.5.3, “Bonding in the IBM BladeCenter” on page 14 for more information.
- It is possible to use VRRP (or equivalent) on the upstream switches as well to provide an even more robust HA design.

6.4.6 L3 configuration with VRRP and static routes

To demonstrate Layer 3 and high availability, the topology in Figure 6-3 on page 114 was used. VRRP has been used for high availability.

Summary of IP addressing used for this configuration

Table 6-1 summarizes the IP addresses used for VLAN 10 and under. VLAN 3 and 10 are used internal to the BladeCenter. VLAN 5 connects from 10GESM1 to C6509 according to the topology map.

Table 6-1 VLAN and IP address info for the switch connections

Switch	VLAN 3	VLAN 5	VLAN 10
10GESM1	10.10.3.1	10.10.5.1	10.10.10.1
10GESM2	10.10.3.2	None	10.10.10.2
C6509	None	10.10.5.200	None
C6506	None	None	None
VRRP - 10GESMs	10.10.3.254	None	10.10.10.254

Table 6-2 summarizes the remaining IP addresses used in the experiment. VLAN 15 connects 10GESM1 to C6506. VLAN 20 connects 10GESM2 to C6506. VLAN 25 connects 10GESM2 to C6509. VLAN 30 connects C6509 to C6506. This is all shown on the topology map.

Table 6-2 VLAN and IP address info for the switch connections

Switch	VLAN 15	VLAN 20	VLAN 25	VLAN 30
10GESM1	10.10.15.1	None	None	None
10GESM2	None	10.10.20.2	10.10.25.2	None
C6509	None	None	10.10.25.200	10.10.30.200
C6506	10.10.15.201	10.10.20.201	None	10.10.30.201
VRRP - 10GESMs	None	None	None	None

L3 full mesh topology with VRRP and static routes

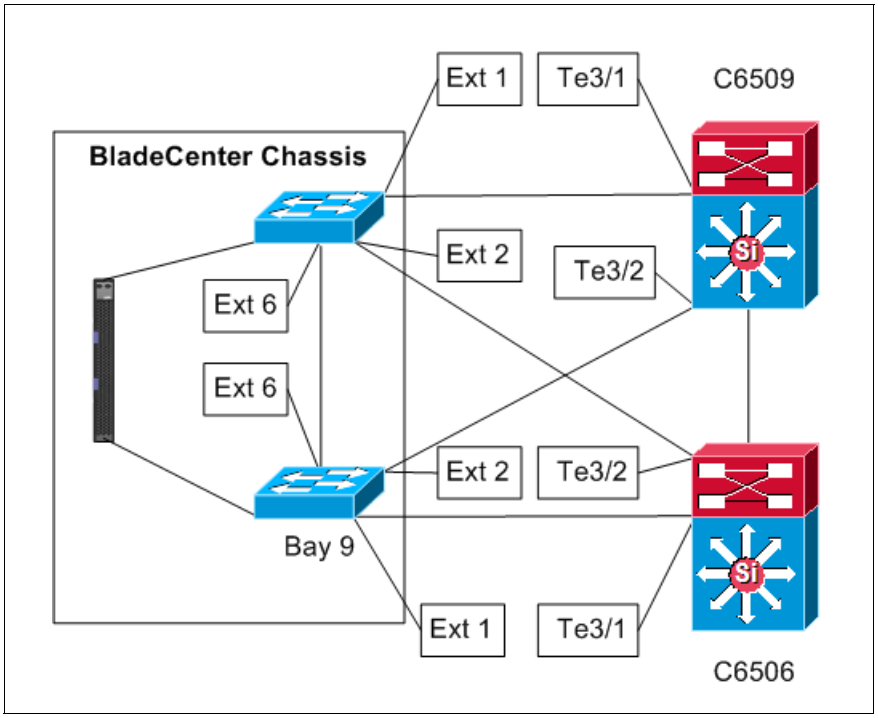


Figure 6-3 L3 full mesh configuration with VRRP and static routes

To complete the topology with static routing, gateways and extra route information were entered manually. The following two examples show gateway information added to the 10GESMs.

Example 6-11 10GESM1: gateways

```
/c/l3/gw 1
        ena
        addr 10.10.15.201
/c/l3/gw 2
        ena
        addr 10.10.5.200
```

Example 6-12 10GESM2: gateways

```
/c/l3/gw 1
        ena
        addr 10.10.25.200
/c/l3/gw 2
        ena
        addr 10.10.20.201
```

The gateway entries on the Nortel 10GESMs reference the two Cisco switches according to the topology map. The next two examples show added route information on both Cisco switches.

Example 6-13 C6509: added routes

```
ip route 0.0.0.0 0.0.0.0 9.42.170.1 (default gateway to external network)
ip route 10.10.3.0 255.255.255.0 10.10.25.2
```

```
ip route 10.10.3.0 255.255.255.0 10.10.5.1
ip route 10.10.10.0 255.255.255.0 10.10.5.1
ip route 10.10.10.0 255.255.255.0 10.10.25.2
```

Example 6-14 C6506: added routes

```
ip route 9.0.0.0 255.0.0.0 10.10.30.200 (default via C6509)
ip route 10.10.3.0 255.255.255.0 10.10.15.1
ip route 10.10.3.0 255.255.255.0 10.10.20.2
ip route 10.10.10.0 255.255.255.0 10.10.15.1
ip route 10.10.10.0 255.255.255.0 10.10.20.2
```

Example 6-15 shows the VRRP settings for both Nortel 10GESMs.

Example 6-15 VRRP settings to enable on both 10GESM1 and 10GESM2

```
/c/l3/vrrp/on
/c/l3/vrrp/vr 3
        ena
        vrid 3
        if 2
        prio 103
        addr 10.10.3.254
/c/l3/vrrp/vr 3/track
                ports e
/c/l3/vrrp/vr 10
        ena
        vrid 10
        if 1
        prio 103
        addr 10.10.10.254
/c/l3/vrrp/vr 10/track
                ports e
```

Note: The **vr** numbers configured in the sample are unique to the switch and are purely local instance numbers. The **vrid** numbers must be the same on all switches which work together and must be unique to that group of switches on the VLAN where they are used.

After setting up VRRP on the Nortel 10GESMs the topology and configuration were verified by pinging all IP addresses. The following two examples show VRRP state information for the Nortel 10GESMs. Note here that 10.10.10.254 is master on 10GESM1.

Example 6-16 10GESM1: VRRP information

```
>> Main# /i/l3/vrrp
VRRP information:
   3: vrid   3, 10.10.3.254,     if  2, renter, prio 107, master
  10: vrid  10, 10.10.10.254,    if  1, renter, prio 113, master
```

Example 6-17 10GESM2: VRRP information

```
>> Main# /i/l3/vrrp
VRRP information:
   3: vrid   3, 10.10.3.254,     if  2, renter, prio 104, backup
  10: vrid  10, 10.10.10.254,    if  1, renter, prio 110, backup
```

During this experiment, a ping will run from a blade server to an IP address on C6509 (10.10.25.200) that is external to the VRRP. This will show the lost transmission during the test. Example 6-18 shows setup and verification of the experiment from the blade server. This time the operating system used is SLES 10.

Example 6-18 Blade server setup and verification

```
bond0     Link encap:Ethernet  HWaddr 00:0E:1E:00:12:2A
          inet addr:10.10.10.101  Bcast:10.10.10.255  Mask:255.255.255.0
          inet6 addr: fe80::20e:1eff:fe00:122a/64 Scope:Link
          UP BROADCAST RUNNING MASTER MULTICAST  MTU:1500  Metric:1
          RX packets:1340 errors:0 dropped:1308 overruns:0 frame:0
          TX packets:40 errors:0 dropped:0 overruns:0 carrier:0
          collisions:0 txqueuelen:0
          RX bytes:82020 (80.0 Kb)  TX bytes:3160 (3.0 Kb)

/* Ping the two Nortel switches (10.10.10.1 and 2) */
hs21-sles10:~ # ping 10.10.10.1
PING 10.10.10.1 (10.10.10.1) 56(84) bytes of data.
64 bytes from 10.10.10.1: icmp_seq=1 ttl=255 time=0.354 ms

--- 10.10.10.1 ping statistics ---
1 packets transmitted, 1 received, 0% packet loss, time 0ms
rtt min/avg/max/mdev = 0.354/0.354/0.354/0.000 ms


hs21-sles10:~ # ping 10.10.10.2
PING 10.10.10.2 (10.10.10.2) 56(84) bytes of data.
64 bytes from 10.10.10.2: icmp_seq=1 ttl=255 time=0.338 ms
64 bytes from 10.10.10.2: icmp_seq=2 ttl=255 time=4.56 ms

--- 10.10.10.2 ping statistics ---
2 packets transmitted, 2 received, 0% packet loss, time 1000ms
rtt min/avg/max/mdev = 0.338/2.453/4.568/2.115 ms


/* add default route to point to the VRRP address */

hs21-sles10:~ # route add -net 10.0.0.0/24 gw 10.10.10.254 metric 1
hs21-sles10:~ # netstat -r
Kernel IP routing table
Destination     Gateway         Genmask         Flags   MSS Window  irtt Iface
10.0.0.0        10.10.10.254    255.255.255.0   UG        0 0          0 bond0
10.10.10.0      *               255.255.255.0   U         0 0          0 bond0
9.42.170.0      *               255.255.254.0   U         0 0          0 eth0
9.42.170.0      *               255.255.254.0   U         0 0          0 eth1
link-local      *               255.255.0.0     U         0 0          0 eth0
loopback        *               255.0.0.0       U         0 0          0 lo
default         9.42.170.1      0.0.0.0         UG        0 0          0 eth0

/* ping the VRRP address */
hs21-sles10:~ # ping 10.10.10.254
PING 10.10.10.254 (10.10.10.254) 56(84) bytes of data.
64 bytes from 10.10.10.254: icmp_seq=1 ttl=255 time=3.22 ms

--- 10.10.10.254 ping statistics ---
```

```
1 packets transmitted, 1 received, 0% packet loss, time 0ms
rtt min/avg/max/mdev = 3.220/3.220/3.220/0.000 ms
```

To test failover with VRRP and trunk failover enabled, external ports on 10GESM1 were disabled from its CLI. Example 6-19 shows the process of failing links from 10GESM1. Both EXT1 and EXT2 were disabled and it can be seen what IP address the switch is trying to contact during that process. After the external ports are failed, the switch automatically disables the internal ports so that blade traffic is now directed to the second 10GESM. In this case, that second switch is 10GESM2. The virtual router becomes a backup after this failover occurs.

Example 6-19 10GESM1 CLI showing failover with VRRP enabled

```
>> Layer 3# /oper/port ext1/dis
Port EXT1 disabled.

>> Operations Port EXT1#

/* console messages show the loss of the connection to the 6509 */
Jan  2  5:05:18 10.10.10.1 NOTICE  system: link down on port EXT1
Jan  2  5:05:18 10.10.10.1 ALERT   ip: cannot contact default gateway 10.10.5.200
Jan  2  5:05:18 10.10.10.1 NOTICE  ip: default gateway 10.10.15.201 enabled

/* This Nortel is still the VRRP master - it still is connected to the 6506 */
/i/l3/vrrp
VRRP information:
   3: vrid   3, 10.10.3.254,     if  2, renter, prio 107, master
  10: vrid  10, 10.10.10.254,    if  1, renter, prio 113, master

>> Layer 3# /oper/port ext2/dis
Port EXT2 disabled.


/* now we trigger failover to the other Nortel switch - and NIC bonding failover
at the same time because the internal ports to the active servers are disabled*/

Jan  2  5:05:36 10.10.10.1 NOTICE  system: link down on port EXT2
Jan  2  5:05:36 10.10.10.1 ALERT   ip: cannot contact default gateway 10.10.15.201
Jan  2  5:05:36 10.10.10.1 NOTICE  server: link down on port INT1
Jan  2  5:05:36 10.10.10.1 NOTICE  server: link down on port INT7
Jan  2  5:05:36 10.10.10.1 NOTICE  server: link down on port INT13
Jan  2  5:05:36 10.10.10.1 NOTICE  server: link down on port INT14
Jan  2  5:05:37 10.10.10.1 NOTICE  vrrp: virtual router    10.10.10.254 is now
BACKUP
```

Example 6-20 shows CLI text indicating that 10.10.10.254 is now master on 10GESM2.

Example 6-20 10GESM2 CLI showing 10.10.10.254 VRRP master

```
>> Layer 3#
Jan  2  0:05:20 10.10.10.2 NOTICE  vrrp: virtual router    10.10.10.254 is now
MASTER
```

A failback was done to let 10GESM1 return to its original state. Example 6-21 on page 118 shows the recovery process.

Example 6-21 10GESM1 CLI recovering the EXT ports

```
>> Operations Port EXT2# /oper/port ext2/ena
Port EXT2 enabled.
>> Operations Port EXT2#

/* internal ports are re-enabled and the switch pre-empts and takes over as VRRP
master */
Jan  2  5:09:49 10.10.10.1 NOTICE  system: link up on port EXT2
Jan  2  5:09:49 10.10.10.1 NOTICE  server: link up on port INT1
Jan  2  5:09:49 10.10.10.1 NOTICE  server: link up on port INT7
Jan  2  5:09:49 10.10.10.1 NOTICE  server: link up on port INT13
Jan  2  5:09:49 10.10.10.1 NOTICE  server: link up on port INT14
Jan  2  5:09:50 10.10.10.1 NOTICE  vrrp: virtual router    10.10.10.254 is now
MASTER
Jan  2  5:09:53 10.10.10.1 NOTICE  ip: default gateway 10.10.15.201 operational
Jan  2  5:09:53 10.10.10.1 NOTICE  ip: default gateway 10.10.15.201 enabled

/* re-enabling the 2nd link just completes the mesh */
/oper/port ext1/ena
Port EXT1 enabled.
>> Operations Port EXT1#
Jan  2  5:10:06 10.10.10.1 NOTICE  system: link up on port EXT1
>> Operations Port EXT1#
Jan  2  5:10:09 10.10.10.1 NOTICE  ip: default gateway 10.10.5.200 operational
```

Example 6-22 shows that 10GESM2 has reverted back to its original state.

Example 6-22 10GESM2 CLI information showing 10.10.10.254 VRRP backup

```
Jan  2  0:09:33 10.10.10.2 NOTICE  vrrp: virtual router    10.10.10.254 is now
BACKUP
```

Example 6-23 shows that 104 of 107 packets were received. This occurred during the initial failure and not during failback. All three were lost between icmp_seq=20 to icmp_seq=24.

Example 6-23 Blade server ping statistics

```
64 bytes from 10.10.25.200: icmp_seq=18 ttl=254 time=0.319 ms
64 bytes from 10.10.25.200: icmp_seq=19 ttl=254 time=0.258 ms
64 bytes from 10.10.25.200: icmp_seq=20 ttl=254 time=0.257 ms
64 bytes from 10.10.25.200: icmp_seq=24 ttl=254 time=0.326 ms
64 bytes from 10.10.25.200: icmp_seq=25 ttl=254 time=0.290 ms
64 bytes from 10.10.25.200: icmp_seq=26 ttl=254 time=0.472 ms


--- 10.10.25.200 ping statistics ---
107 packets transmitted, 104 received, 2% packet loss, time 106057ms
rtt min/avg/max/mdev = 0.248/0.323/0.673/0.072 ms
```

Full configuration dumps are located in A.1.3, “Configurations for 6.4.6, “L3 configuration with VRRP and static routes” on page 113” on page 157.

6.4.7 L3 configuration with VRRP and RIP

The topology from the previous section was modified to include RIP. The static routes were pulled out of the switch configurations and RIP was enabled.

Summary of IP Addressing used for this configuration

Table 6-3 summarizes the IP addresses used for VLAN 10 and under. VLAN 3 and 10 are used internal to the BladeCenter. VLAN 5 connects from 10GESM1 to C6509 according to the topology map.

Table 6-3 VLAN and IP address info for the switch connections

Switch	VLAN 1	VLAN 3	VLAN 5	VLAN 10
10GESM1	9.42.170.246	10.10.3.1	10.10.5.1	10.10.10.1
10GESM2	9.42.170.247	10.10.3.2	None	10.10.10.2
C6509	9.42.170.243	None	10.10.5.200	None
C6506	None	None	None	None
VRRP - 10GESMs	None	10.10.3.254	None	10.10.10.254

Table 6-4 summarizes the remaining IP addresses used in the experiment. VLAN 15 connects 10GESM1 to C6506. VLAN 20 connects 10GESM2 to C6506. VLAN 25 connects 10GESM2 to C6509. VLAN 30 connects C6509 to C6506. This is all corresponding to the topology map.

Table 6-4 VLAN and IP address info for the switch connections

Switch	VLAN 15	VLAN 20	VLAN 25	VLAN 30
10GESM1	10.10.15.1	None	None	None
10GESM2	None	10.10.20.2	10.10.25.2	None
C6509	None	None	10.10.25.200	10.10.30.200
C6506	10.10.15.201	10.10.20.201	None	10.10.30.201
VRRP - 10GESMs	None	None	None	None

L3 full mesh topology with VRRP and RIP

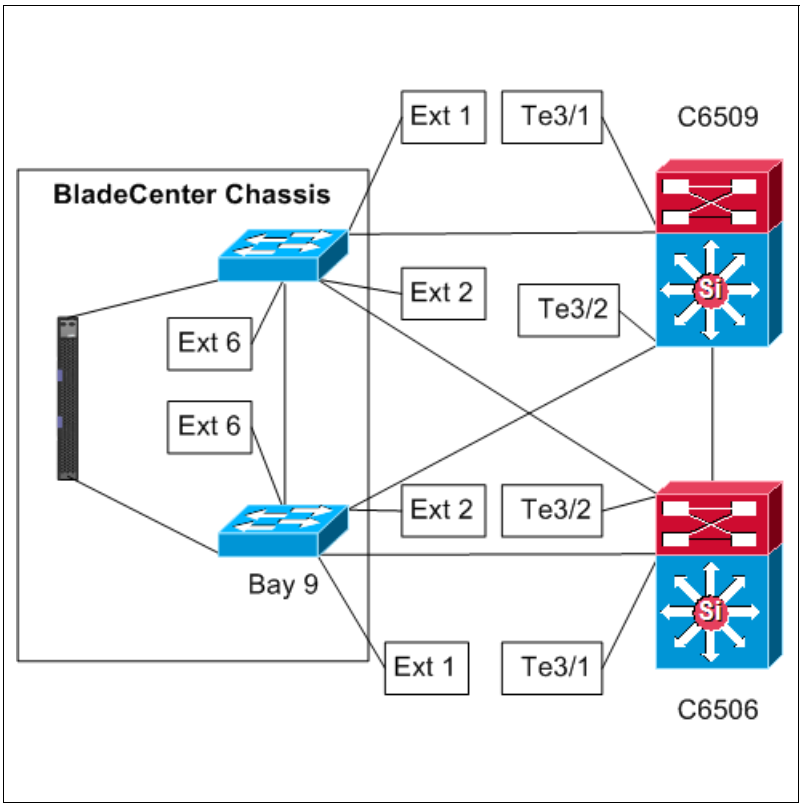


Figure 6-4 L3 full mesh configuration with VRRP and RIP

Figure 6-4 shows the evolving topology map. This configuration was taken up to dynamic routing starting out with RIP version 2.

Example 6-24 shows the modifications to the two Cisco switches. The old static routes were taken out. Those static routes are not needed once all the switches are sharing route information with RIP.

Example 6-24 C6509 and C6506 configuration additions for RIP

```
router rip
 version 2
 network 10.0.0.0
```

Example 6-25 details the configuration changes on 10GESM1 to enable dynamic routing with RIP. Note that most of the configuration text is automatically generated and represents default values; all that typically needs to be entered is the command to enable RIP on each IP interface. The following two examples detail the configuration changes on 10GESM1 and 10GESM2 to enable dynamic routing with RIP.

Example 6-25 10GESM1 configuration additions for RIP

```
/c/l3/rip/updat 30/on
/c/l3/rip/if 1/ena/supply e/listen e/default none/version 2
/c/l3/rip/if 1/poison d/split e/trigg e/metric 1/mcast e
/c/l3/rip/if 1/auth none
/c/l3/rip/if 2/ena/supply e/listen e/default none/version 2
/c/l3/rip/if 2/poison d/split e/trigg e/metric 1/mcast e
```

```
/c/l3/rip/if 2/auth none
/c/l3/rip/if 5/ena/supply e/listen e/default none/version 2
/c/l3/rip/if 5/poison d/split e/trigg e/metric 1/mcast e
/c/l3/rip/if 5/auth none
/c/l3/rip/if 15/ena/supply e/listen e/default none/version 2
/c/l3/rip/if 15/poison d/split e/trigg e/metric 1/mcast e
/c/l3/rip/if 15/auth none
```

Example 6-26 10GESM2 configuration additions for RIP

```
/c/l3/rip/updat 30/on
/c/l3/rip/if 1/ena/supply e/listen e/default none/version 2
/c/l3/rip/if 1/poison d/split e/trigg e/metric 1/mcast e
/c/l3/rip/if 1/auth none
/c/l3/rip/if 2/ena/supply e/listen e/default none/version 2
/c/l3/rip/if 2/poison d/split e/trigg e/metric 1/mcast e
/c/l3/rip/if 2/auth none
/c/l3/rip/if 20/ena/supply e/listen e/default none/version 2
/c/l3/rip/if 20/poison d/split e/trigg e/metric 1/mcast e
/c/l3/rip/if 20/auth none
/c/l3/rip/if 25/ena/supply e/listen e/default none/version 2
/c/l3/rip/if 25/poison d/split e/trigg e/metric 1/mcast e
/c/l3/rip/if 25/auth none
```

To verify this configuration the default gateways on 10GESM1 and 10GESM2 were disabled. Example 6-27 shows this being done.

Example 6-27 Disable gateways on 10GESM1 and 10GESM2

```
/c/l3/gw 1
        dis
/c/l3/gw 2
        dis
```

Once this was completed you could see RIP routes added with **/i/l2/route/d**. This was tested by PINGs from the Nortel switches and from blade servers to destinations on or beyond the Cisco switches which involved L3 routing on the Nortel switches.

Full configuration dumps are located in A.1.4, “Configurations for 6.4.7, “L3 configuration with VRRP and RIP” on page 119” on page 168.

6.4.8 L3 configuration with VRRP and OSPF

The topology from the previous section was modified to include OSPF. The entries for RIP were removed and OSPF was configured for this experiment’s dynamic routing.

Summary of IP Addressing used for this configuration

Table 6-1 on page 113 summarizes the IP addresses used for VLAN 10 and under. VLAN 1 is the management VLAN. VLAN 3 and 10 are used internal to the BladeCenter. VLAN 5 connects from 10GESM1 to C6509 according to the topology map.

Table 6-5 VLAN and IP address info for the switch connections

Switch	VLAN 3	VLAN 5	VLAN 10
10GESM1	10.10.3.1	10.10.5.1	10.10.10.1
10GESM2	10.10.3.2	None	10.10.10.2
C6509	None	10.10.5.200	None
C6506	None	None	None
VRRP - 10GESMs	10.10.3.254	None	10.10.10.254

Table 6-2 on page 113 summarizes the remaining IP addresses used in the experiment. VLAN 15 connects 10GESM1 to C6506. VLAN 20 connects 10GESM2 to C6506. VLAN 25 connects 10GESM2 to C6509. VLAN 30 connects C6509 to C6506. This is all corresponding to the topology map.

Table 6-6 VLAN and IP address info for the switch connections

Switch	VLAN 15	VLAN 20	VLAN 25	VLAN 30
10GESM1	10.10.15.1	None	None	None
10GESM2	None	10.10.20.2	10.10.25.2	None
C6509	None	None	10.10.25.200	10.10.30.200
C6506	10.10.15.201	10.10.20.201	None	10.10.30.201
VRRP - 10GESMs	None	None	None	None

L3 full mesh topology with VRRP and OSPF

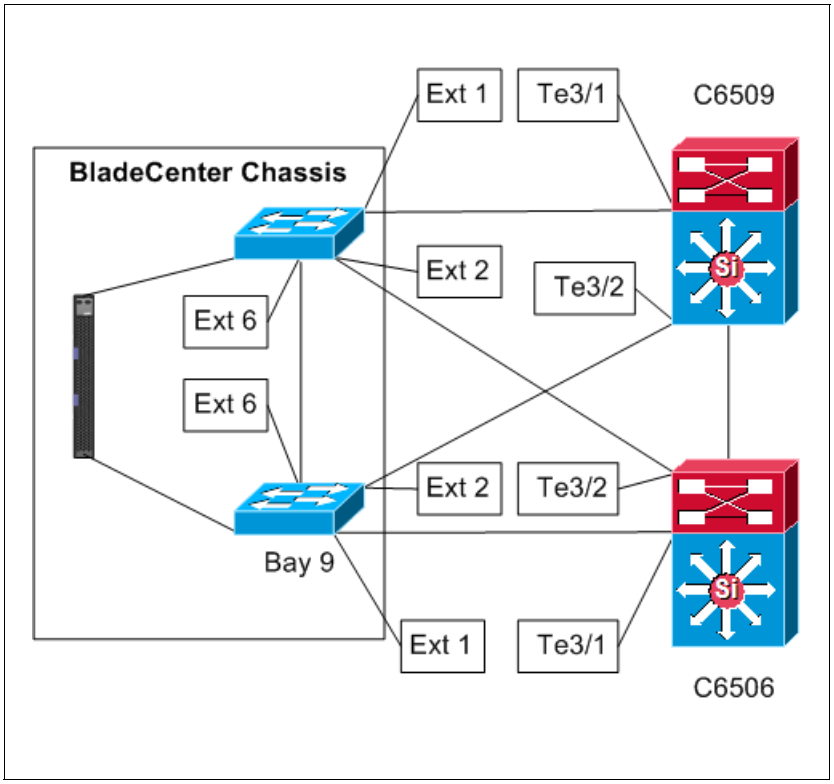


Figure 6-5 L3 full mesh configuration with VRRP and OSPF

Figure 6-5 shows the evolving topology map. The previous dynamic routing configuration using RIP was converted to OSPF.

Example 6-28 shows the modifications to the two Cisco switches. There are no static routes and the RIP modifications from the previous section are removed.

Example 6-28 C6509 and C6506 configuration additions for OSPF

```
router ospf 1
 log-adjacency-changes
 redistribute static subnets
network 10.0.0.0 0.255.255.255 area 0
 default-information originate always
```

Note: The 6506 would not learn the default gateway from the 6509 until the “default originate” command was removed from the 6506. The core switches, not the lower layer ones, should be the ones to originate default routes.

The following two examples detail the configuration changes on 10GESM1 and 10GESM2 to take this topology to OSPF. The entries for RIP were disabled. There are no static routes or default gateways enabled. Most of this configuration is automatically generated when a particular OSPF interface is enabled. It represents the default values except where specifically noted; all that typically needs to be entered is the command to enable ospf on the specified ip interface(s).

Example 6-29 10GESM1 configuration additions for OSPF

```
/c/l3/ospf/if 5
        ena
        aindex 0
        prio 1
        cost 1
        hello 10
        dead 40
        trans 1
        retra 5
/c/l3/ospf/if 15
        ena
        aindex 0
        prio 1
        cost 1
        hello 10
        dead 40
        trans 1
        retra 5
```

Example 6-30 10GESM2 configuration additions for OSPF

```
/c/l3/ospf/if 20
        ena
        aindex 0
        prio 1
        cost 1
        hello 10
        dead 40
        trans 1
        retra 5
/c/l3/ospf/if 25
        ena
        aindex 0
        prio 1
        cost 1
        hello 10
        dead 40
        trans 1
        retra 5
```

No failover test was conducted in this section because it was completed successfully at Layer 3 while using RIP. Routing table dumps were done to verify the interoperability between the Nortel and Cisco switches running OSPF. The following examples detail this process.

Example 6-31 details the findings on C6509. The lines beginning with “O” represent routes learned through OSPF.

Example 6-31 C6509 routing information

```
c6509_A#sh ip route
Codes: C - connected, S - static, R - RIP, M - mobile, B - BGP
       D - EIGRP, EX - EIGRP external, O - OSPF, IA - OSPF inter area
       N1 - OSPF NSSA external type 1, N2 - OSPF NSSA external type 2
       E1 - OSPF external type 1, E2 - OSPF external type 2, E - EGP
```

```
       i - IS-IS, su - IS-IS summary, L1 - IS-IS level-1, L2 - IS-IS level-2
       ia - IS-IS inter area, * - candidate default, U - per-user static route
       o - ODR, P - periodic downloaded static route

Gateway of last resort is 9.42.170.1 to network 0.0.0.0

     9.0.0.0/23 is subnetted, 1 subnets
C       9.42.170.0 is directly connected, Vlan1
     10.0.0.0/24 is subnetted, 7 subnets
O       10.10.3.0 [110/2] via 10.10.25.2, 01:26:31, TenGigabitEthernet3/2
                  [110/2] via 10.10.5.1, 01:26:31, TenGigabitEthernet3/1
C       10.10.5.0 is directly connected, TenGigabitEthernet3/1
O       10.10.10.0 [110/2] via 10.10.25.2, 01:26:31, TenGigabitEthernet3/2
                   [110/2] via 10.10.5.1, 01:26:31, TenGigabitEthernet3/1
O       10.10.15.0 [110/2] via 10.10.30.201, 01:26:31, Vlan30
                   [110/2] via 10.10.5.1, 01:26:31, TenGigabitEthernet3/1
O       10.10.20.0 [110/2] via 10.10.30.201, 01:26:31, Vlan30
                   [110/2] via 10.10.25.2, 01:26:31, TenGigabitEthernet3/2
C       10.10.25.0 is directly connected, TenGigabitEthernet3/2
C       10.10.30.0 is directly connected, Vlan30
S*   0.0.0.0/0 [1/0] via 9.42.170.1

c6509_A#sh ip ospf neigh

Neighbor ID     Pri   State           Dead Time   Address         Interface
10.10.3.2         1   FULL/BDR        00:00:32    10.10.25.2
TenGigabitEthernet3/2
10.10.3.1         1   FULL/BDR        00:00:35    10.10.5.1
TenGigabitEthernet3/1
10.10.30.201      1   FULL/DR         00:00:38    10.10.30.201    Vlan30
```

Example 6-32 details the findings on C6506.

Example 6-32 C6506 routing information

```
c6506#sh ip route
Codes: C - connected, S - static, R - RIP, M - mobile, B - BGP
       D - EIGRP, EX - EIGRP external, O - OSPF, IA - OSPF inter area
       N1 - OSPF NSSA external type 1, N2 - OSPF NSSA external type 2
       E1 - OSPF external type 1, E2 - OSPF external type 2, E - EGP
       i - IS-IS, su - IS-IS summary, L1 - IS-IS level-1, L2 - IS-IS level-2
       ia - IS-IS inter area, * - candidate default, U - per-user static route
       o - ODR, P - periodic downloaded static route

Gateway of last resort is 10.10.30.200 to network 0.0.0.0

     9.0.0.0/23 is subnetted, 1 subnets
O       9.42.170.0 [110/2] via 10.10.30.200, 01:28:25, Vlan30
     10.0.0.0/24 is subnetted, 7 subnets
O       10.10.3.0 [110/2] via 10.10.20.2, 01:28:25, Vlan20
                  [110/2] via 10.10.15.1, 01:28:25, Vlan15
O       10.10.5.0 [110/2] via 10.10.30.200, 01:28:25, Vlan30
                  [110/2] via 10.10.15.1, 01:28:25, Vlan15
O       10.10.10.0 [110/2] via 10.10.20.2, 01:28:25, Vlan20
                   [110/2] via 10.10.15.1, 01:28:25, Vlan15
C       10.10.15.0 is directly connected, Vlan15
```

```
C       10.10.20.0 is directly connected, Vlan20
0       10.10.25.0 [110/2] via 10.10.20.2, 01:28:25, Vlan20
                   [110/2] via 10.10.30.200, 01:28:25, Vlan30
C       10.10.30.0 is directly connected, Vlan30
O*E2 0.0.0.0/0 [110/1] via 10.10.30.200, 01:28:28, Vlan30

c6506#sh ip ospf neigh

Neighbor ID     Pri   State           Dead Time   Address         Interface
10.10.30.200      1   FULL/BDR        00:00:38    10.10.30.200    Vlan30
10.10.3.2         1   FULL/BDR        00:00:32    10.10.20.2      Vlan20
10.10.3.1         1   FULL/BDR        00:00:38    10.10.15.1      Vlan15
```

Example 6-33 details the findings on 10GESM1.

Example 6-33 10GESM1 routing information

```
>> Configuration# /i/l3/ospf/nbr
-------------------------------------------------------------
Intf  NeighborID       Prio  State         Address
----  ----------       ----  -----         -------
   1  10.10.3.2           1  Full           10.10.10.2
   2  10.10.3.2           1  Full           10.10.3.2
   5  10.10.30.200        1  Full           10.10.5.200
  15  10.10.30.201        1  Full           10.10.15.201

>> OSPF Database# /i/l3/ip
IP information:
  Router ID: 10.10.3.1,  AS number 0

Interface information:
  1: 10.10.10.1      255.255.255.0   10.10.10.255,    vlan 10, up
  2: 10.10.3.1       255.255.255.0   10.10.3.255,     vlan 3, up
  5: 10.10.5.1       255.255.255.0   10.10.5.255,     vlan 5, up
 15: 10.10.15.1      255.255.255.0   10.10.15.255,    vlan 15, up
250: 9.42.170.246    255.255.254.0   9.42.171.255,    vlan 4095, up

Default gateway information: metric strict
  1: 10.10.15.201,    disabled
  2: 10.10.5.200,     disabled
254: 9.42.170.1,      up  active

Current BOOTP relay settings: OFF
  0.0.0.0, 0.0.0.0

Current IP forwarding settings: ON, dirbr disabled


Current network filter settings:
  none

Current route map settings:

Current OSPF settings: ON
  Default route none
  Router ID: 10.10.3.1
```

```
  lsdb limit 0

Current OSPF area settings:
  0: 0.0.0.0,         type transit, auth none, metric 1, spf 10, enabled

Current OSPF interface settings:
  1: 10.10.10.1,      area index  0, prio 1, cost 1, enabled
    hello 10, dead 40, trans 1, retra 5
    key
  2: 10.10.3.1,       area index  0, prio 1, cost 1, enabled
    hello 10, dead 40, trans 1, retra 5
    key
  5: 10.10.5.1,       area index  0, prio 1, cost 1, enabled
    hello 10, dead 40, trans 1, retra 5
    key
 15: 10.10.15.1,      area index  0, prio 1, cost 1, enabled
    hello 10, dead 40, trans 1, retra 5
    key

>> Layer 3# /i/l3/route
------------------------------------------------------------
[IP Routing Menu]
     find     - Show a single route by destination IP address
     gw       - Show routes to a single gateway
     type     - Show routes of a single type
     tag      - Show routes of a single tag
     if       - Show routes on a single interface
     dump     - Show all routes

>> IP Routing# du
Status code: * - best
    Destination         Mask          Gateway        Type      Tag     Metr If
  --------------- --------------- --------------- --------- --------- ---- --
  0.0.0.0         0.0.0.0         10.10.5.200     indirect  ospf         1  5
* 9.42.170.0      255.255.254.0   10.10.5.200     indirect  ospf         2  5
* 10.10.3.0       255.255.255.0   10.10.3.1       direct    fixed           2
* 10.10.3.1       255.255.255.255 10.10.3.1       local     addr            2
* 10.10.3.255     255.255.255.255 10.10.3.255     broadcast broadcast       2
* 10.10.5.0       255.255.255.0   10.10.5.1       direct    fixed           5
* 10.10.5.1       255.255.255.255 10.10.5.1       local     addr            5
* 10.10.5.255     255.255.255.255 10.10.5.255     broadcast broadcast       5
* 10.10.10.0      255.255.255.0   10.10.10.1      direct    fixed           1
* 10.10.10.1      255.255.255.255 10.10.10.1      local     addr            1
* 10.10.10.255    255.255.255.255 10.10.10.255    broadcast broadcast       1
* 10.10.15.0      255.255.255.0   10.10.15.1      direct    fixed          15
* 10.10.15.1      255.255.255.255 10.10.15.1      local     addr           15
* 10.10.15.255    255.255.255.255 10.10.15.255    broadcast broadcast      15
* 10.10.20.0      255.255.255.0   10.10.10.2      indirect  ospf         2  1
* 10.10.25.0      255.255.255.0   10.10.5.200     indirect  ospf         2  5
* 10.10.30.0      255.255.255.0   10.10.5.200     indirect  ospf         2  5
* 127.0.0.0       255.0.0.0       0.0.0.0         martian   martian
* 224.0.0.0       224.0.0.0       0.0.0.0         martian   martian
* 224.0.0.0       240.0.0.0       0.0.0.0         multicast addr
* 224.0.0.2       255.255.255.255 0.0.0.0         multicast addr
* 224.0.0.5       255.255.255.255 0.0.0.0         multicast addr
```

```
* 224.0.0.6       255.255.255.255 0.0.0.0         multicast addr
* 224.0.0.18      255.255.255.255 0.0.0.0         multicast addr
* 255.255.255.255 255.255.255.255 255.255.255.255 broadcast broadcast
```

Example 6-34 details the findings on 10GESM2.

Example 6-34 10GESM2 routing information

```
>> Configuration# /i/l3/ip
IP information:
  Router ID: 10.10.3.2,  AS number 0

Interface information:
  1: 10.10.10.2      255.255.255.0   10.10.10.255,    vlan 10, up
  2: 10.10.3.2       255.255.255.0   10.10.3.255,     vlan 3, up
 20: 10.10.20.2      255.255.255.0   10.10.20.255,    vlan 20, up
 25: 10.10.25.2      255.255.255.0   10.10.25.255,    vlan 25, up
250: 9.42.170.247    255.255.254.0   9.42.171.255,    vlan 4095, up

Default gateway information: metric strict
  1: 10.10.25.200,    disabled
  2: 10.10.20.201,    disabled
254: 9.42.170.1,      up  active

Current BOOTP relay settings: OFF
  0.0.0.0, 0.0.0.0

Current IP forwarding settings: ON, dirbr disabled


Current network filter settings:
  none

Current route map settings:

Current OSPF settings: ON
  Default route none
  Router ID: 10.10.3.2
  lsdb limit 0

Current OSPF area settings:
  0: 0.0.0.0,         type transit, auth none, metric 1, spf 10, enabled

Current OSPF interface settings:
  1: 10.10.10.2,      area index  0, prio 1, cost 1, enabled
    hello 10, dead 40, trans 1, retra 5
    key
  2: 10.10.3.2,       area index  0, prio 1, cost 1, enabled
    hello 10, dead 40, trans 1, retra 5
    key
 20: 10.10.20.2,      area index  0, prio 1, cost 1, enabled
    hello 10, dead 40, trans 1, retra 5
    key
 25: 10.10.25.2,      area index  0, prio 1, cost 1, enabled
    hello 10, dead 40, trans 1, retra 5
    key
```

```
>> Layer 3# /i/l3/ospf/nbr
Intf  NeighborID       Prio  State         Address
----  ----------       ----  -----         -------
   1  10.10.3.1           1  Full           10.10.10.1
   2  10.10.3.1           1  Full           10.10.3.1
  20  10.10.30.201        1  Full           10.10.20.201
  25  10.10.30.200        1  Full           10.10.25.200

# /i/l3/route/d
Status code: * - best
    Destination         Mask          Gateway        Type      Tag     Metr If
  --------------- --------------- --------------- --------- --------- ---- --
  0.0.0.0         0.0.0.0         10.10.25.200    indirect  ospf         1 25
* 9.42.170.0      255.255.254.0   10.10.25.200    indirect  ospf         2 25
* 10.10.3.0       255.255.255.0   10.10.3.2       direct    fixed          2
* 10.10.3.2       255.255.255.255 10.10.3.2       local     addr           2
* 10.10.3.255     255.255.255.255 10.10.3.255     broadcast broadcast      2
* 10.10.5.0       255.255.255.0   10.10.3.1       indirect  ospf         2  2
* 10.10.10.0      255.255.255.0   10.10.10.2      direct    fixed          1
* 10.10.10.2      255.255.255.255 10.10.10.2      local     addr           1
* 10.10.10.255    255.255.255.255 10.10.10.255    broadcast broadcast      1
* 10.10.20.0      255.255.255.0   10.10.20.2      direct    fixed         20
* 10.10.20.2      255.255.255.255 10.10.20.2      local     addr          20
* 10.10.20.255    255.255.255.255 10.10.20.255    broadcast broadcast     20
* 10.10.25.0      255.255.255.0   10.10.25.2      direct    fixed         25
* 10.10.25.2      255.255.255.255 10.10.25.2      local     addr          25
* 10.10.25.255    255.255.255.255 10.10.25.255    broadcast broadcast     25
* 10.10.30.0      255.255.255.0   10.10.20.201    indirect  ospf         2 20
* 127.0.0.0       255.0.0.0       0.0.0.0         martian   martian
* 224.0.0.0       224.0.0.0       0.0.0.0         martian   martian
* 224.0.0.0       240.0.0.0       0.0.0.0         multicast addr
* 224.0.0.2       255.255.255.255 0.0.0.0         multicast addr
* 224.0.0.5       255.255.255.255 0.0.0.0         multicast addr
* 224.0.0.6       255.255.255.255 0.0.0.0         multicast addr
* 224.0.0.18      255.255.255.255 0.0.0.0         multicast addr
* 255.255.255.255 255.255.255.255 255.255.255.255 broadcast broadcast
```

6.5 Stacking configuration with 10 Gb Uplink switches

The configuration and topology in this section use the 10 Gb Uplink switches. This topology shows a “chassis stacking” technique which allows multiple chassis to share one or more 10 Gb connections to the core network. It is likely to be used where the bandwidth requirements of the applications running in multiple BladeCenter chassis can be satisfied by a smaller number of links and where it is desired to avoid the cost of 10 Gb ports on the core switches. Of course this design can be upgraded to add additional 10 Gb links if the bandwidth requirements should grow.

Note: This configuration can be oversubscribed, but it should be noted that the Nortel 10 Gb Uplink switch module has been tested and is capable of supporting concurrent wire speed traffic on all of its ports. Detailed information on this testing can be found at: *http://www.tolly.com/DocDetail.aspx?DocNumber=206168*

The physical topology tested was as follows:

- Three Nortel 10 Gb Uplink switch modules
- Each switch module was connected to the other two at 10 Gb using CX4 copper cable.
- One of the three switches was connected to a Cisco 6509 at 10 Gb using 10 GbaseSR fiber, which required an XFP transceiver on the Nortel switch module.

This topology is shown in Figure 6-6.

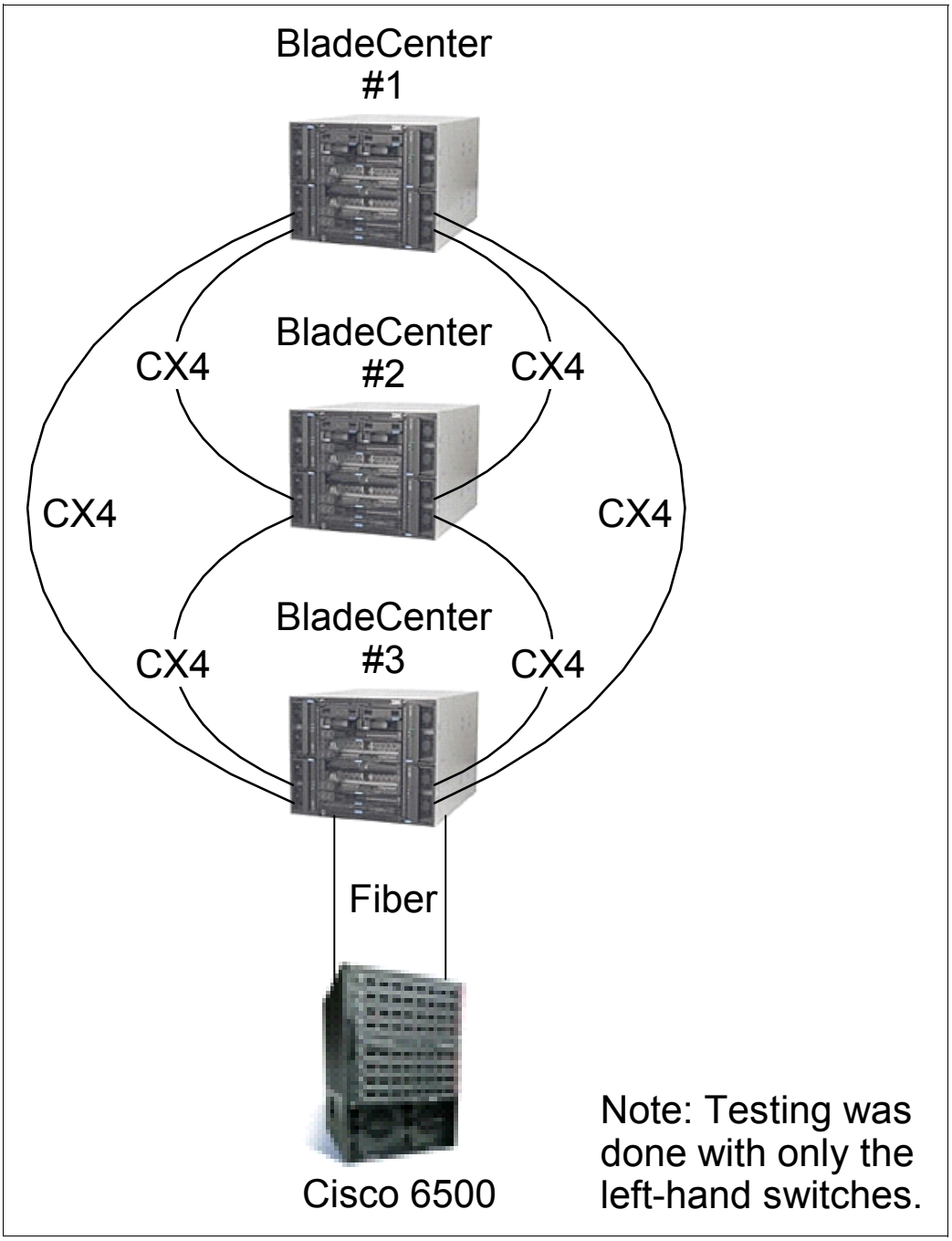


Figure 6-6 Physical topology for chassis stacking with Nortel 10 Gb uplink switches

Figure 6-6 on page 130 shows two switches per chassis. To provide redundancy, two switches per chassis would be deployed and configured to support NIC teaming, trunk failover, and possibly VRRP using configurations of the type shown in Sections 6.3, “Layer 2 configurations” on page 101 and 6.4, “Layer 3 configurations” on page 109. The testing of chassis stacking was done with only one switch per chassis due to hardware limitations.

Two configurations were tested with this topology, one at Layer 2 and one at Layer 3.

6.5.1 Layer 2 configuration

This configuration is modeled after the recommendation shown in the note at the end of Section 6.3.5, “Bonding and Teaming in a Layer 2 topology” on page 107. It uses MSTP on the Cisco switch and PVST with *Uplink Fast* on the Nortel switches.

Uplink Fast is used rather than *Fast Forwarding* because the Nortel switch modules are connected to each other in a physical loop in this topology and their ports are not truly edge ports in the same sense as in other examples. The functionality provided by *Uplink Fast* enables one of the Nortel 10 Gb Uplink switches which has a blocked port leading to one other 10 Gb Uplink switch to fail over to the blocked port immediately if its current active port (which will always be a root port - the path to the root switch) should fail. Like *Fast Forward*, this avoids the up to 50 second delay typically encountered when using classic STP.

The messages and process of failover with *Uplink Fast* is illustrated in Example 6-35. The Spanning Tree status of the Internal ports was deleted from the messages to save space, because STP is disabled on all of the Internal ports throughout this test.

In Example 6-35, initially port EXT1 is the root port and is carrying all traffic to the other two Nortel 10 Gb Switch Modules as well as to the remainder of the network through the Cisco 6509. When EXT1 is disabled, the switch immediately sets port EXT2 to FWD (forwarding) status, and it becomes the root port, even though EXT1 and EXT2 are connected to two different switches.

Example 6-35 Uplink Fast with Layer 2 Chassis Stacking configuration

```
Layer 2# /i/l2/stg 1
------------------------------------------------------------------

upfast enabled, update 40
------------------------------------------------------------------
Spanning Tree Group 1: On (STP/PVST+)
VLANs:  1

Current Root:            Path-Cost  Port Hello MaxAge FwdDel Aging
 6000 00:0f:f8:8c:78:00     6004    EXT1    2     20     15    300

Parameters:  Priority  Hello  MaxAge  FwdDel  Aging
              65500      2      20      15     300

Port  Priority   Cost   FastFwd     State      Designated Bridge     Des Port
----  --------   ----   --------   -------     --------------------  --------
EXT1       128   3002!      n    FORWARDING    ffdc-00:16:ca:ff:48:00    8012
EXT2       128   3002!      n     BLOCKING     ffdc-00:16:ca:ff:06:00    8012
EXT3       128      0!      n     DISABLED
EXT4       128      0!      n     DISABLED
* = STP turned off for this port.
! = Automatic path cost.
```

```
>> Layer 2# /oper/port ext1/dis
Port EXT1 disabled.

>> Operations Port EXT1#
Jan  1 20:02:01 9.42.170.240 ALERT   stg: STG 1 root port EXT1 has gone down. P
tting backup Fast Uplink port EXT2 into forwarding

Jan  1 20:02:01 9.42.170.240 NOTICE  system: link down on port EXT1

>> Operations Port EXT1# /i/l2/stg 1
------------------------------------------------------------------

upfast enabled, update 40

------------------------------------------------------------------
Spanning Tree Group 1: On (STP/PVST+)
VLANs:  1

Current Root:            Path-Cost  Port Hello MaxAge FwdDel Aging
 6000 00:0f:f8:8c:78:00     9006    EXT2    2     20     15    300

Parameters:  Priority  Hello  MaxAge  FwdDel  Aging
              65500      2      20      15     300

Port  Priority   Cost   FastFwd     State      Designated Bridge     Des Port
----  --------   ----   --------   -------     --------------------  --------
EXT1       128   3002!      n     DISABLED
EXT2       128   3002!      n    FORWARDING    ffdc-00:16:ca:ff:06:00    8012
EXT3       128      0!      n     DISABLED
EXT4       128      0!      n     DISABLED
* = STP turned off for this port.
! = Automatic path cost.

>> Layer 2# /oper/port ext1/ena
Port EXT1 enabled.

>> Operations Port EXT1#
Jan  1 20:02:53 9.42.170.240 NOTICE  system: link up on port EXT1

Jan  1 20:02:54 9.42.170.240 ALERT   stg: STG 1 preferred Fast Uplink port EXT1
active. Waiting 35 seconds before switching from port EXT2

Jan  1 20:02:54 9.42.170.240 ALERT   stg: Port EXT1, putting port into blocking
state

Jan  1 20:03:28 9.42.170.240 ALERT   stg: Setting STG 1 Fast Uplink primary por
 EXT1 forwarding and backup port EXT2 blocking
```

The Cisco 6509 sees the port connecting to the Nortel 10 Gb Uplink switch module (only one is connected) as a boundary port. If only one STP instance is in use it is an STP boundary; if we activate a second STP instance (not necessary even if multiple VLANs are in use) then it is a PVST boundary.

The configurations on the three Nortel switches in this test are identical and are shown in Example 6-36. The only line that needed to be entered is the one that enables *Uplink Fast*.

Example 6-36 Nortel Configuration for Uplink Fast

```
>> Main# /c/d
script start "Nortel 10Gb Uplink Ethernet Switch Module for IBM BladeCenter" 5
/**** DO NOT EDIT THIS LINE!
/* Configuration dump taken 20:13:34 Thu Jan  1, 2070
/* Version 1.0.1.1,  Base MAC address 00:16:ca:ff:7a:00
/c/l2/upfast ena/update 40
/c/l2/stg 1/brg/prior 65500
/c/l2/stg 2/brg/prior 65500
.... automatically generated for all of the permitted STG groups ...
/c/l2/stg 31/brg/prior 65500
/
script end  /**** DO NOT EDIT THIS LINE!
```

6.5.2 Layer 3 Configuration

A configuration using OSPF was tested using this topology. A sample of the configuration commands used on the switches are shown below (with default configuration commands removed; only what would need to be typed). The three switches' configurations are essentially the same except for the VLAN numbers for the external ports and those needed to create the triangle of three Nortel 10 Gb Uplink Switch Modules and a single link to the Cisco 6500. The VLANs used for the servers are also different on each of the three switches. The logical topology of this test is shown in Figure 6-7 on page 134.

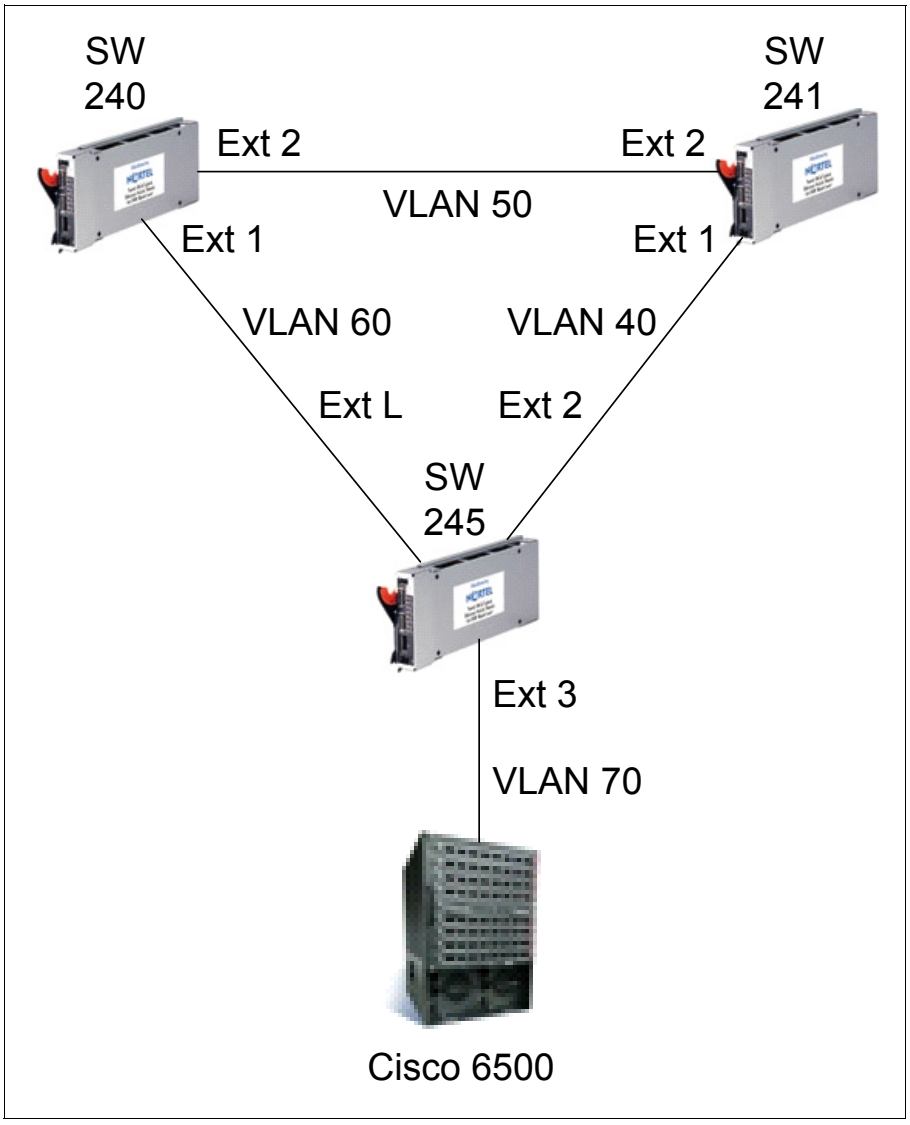


Figure 6-7 Logical topology for Layer 3 chassis stacking

In Figure 6-7, this configuration would also be replicated for the second triad of switches and the pair of switches in each chassis would have additional configuration commands to implement one of the Layer 2 or Layer 3 configurations above. We were not able to actually tests this because it would have required a total of 6 10 Gb Uplink switch modules.

Note: The Layer 3 configuration tested assumes that the servers in the three chassis are all on different VLANs and different subnets. It is not possible to use the same VLAN and subnet across multiple chassis in this configuration because it would cause a split or *discontiguous* subnet. This implies that there are different paths to get to different portions of the same subnet, which violates the IP protocol. If it is desired to have servers in different chassis on the same subnet(s), then the associated VLAN(s) must be connected across those chassis using Layer 2.

Example 6-37 on page 135 shows the configuration modifications.

Example 6-37 Nortel L3 configuration for chassis stacking

```
script start "Nortel 10Gb Uplink Ethernet Switch Module for IBM BladeCenter" 5
/**** DO NOT EDIT THIS LINE!
/* Configuration dump taken 22:20:53 Thu Jan  1, 2070
/* Version 1.0.1.1,  Base MAC address 00:16:ca:ff:7a:00
/c/port EXT1
   pvid 60
/c/port EXT2
   pvid 50
/c/l2/vlan 50
   ena
   name "VLAN 50"
   def EXT2
/c/l2/vlan 60
   ena
   name "VLAN 60"
   def EXT1
/c/l2/vlan 120
   ena
   name "VLAN 120"
   def INT9 INT10 INT11 INT12
/c/l3/ospf/on
/c/l3/ospf/aindex 0
   ena
   areaid 0.0.0.0
/c/l3/ospf/if 50
   ena
/c/l3/ospf/if 60
   ena
/c/l3/ospf/if 120 /* this vlan is where the servers would be */
   ena
   aindex 0
   /c/l3/if 50
   ena
   addr 10.10.50.240
   mask 255.255.255.0
   broad 10.10.50.255
   vlan 50
/c/l3/if 60
   ena
   addr 10.10.60.240
   mask 255.255.255.0
   broad 10.10.60.255
   vlan 60
/c/l3/if 120
   ena
   addr 10.10.120.240
   mask 255.255.255.0
   broad 10.10.120.255
   vlan 120
/
script end  /**** DO NOT EDIT THIS LINE!
```

A

Switch configuration files

This appendix includes switch configuration logs used in networking experiments.

A.1 Switch configuration files used in networking experiments

Included here are the full switch configuration logs that were used in Chapter 6, “Nortel 10 Gb Ethernet Switch Module configuration and network integration” on page 97.

A.1.1 Configurations for 6.3.3, “Layer 2 configuration with MSTP” on page 104

The following examples show switch configuration dumps from the topology using VLAN tagging, static trunk, LACP, and MSTP.

Example: A-1 10GESM1

```
>> Main# /c/d
script start "Nortel 10Gb Ethernet Switch Module for IBM BladeCenter" 5  /***
DO NOT EDIT THIS LINE!
/* Configuration dump taken 13:46:35 Wed May 16, 2007
/* Version 1.0.1.1,  Base MAC address 00:17:ef:71:e0:00
/c/sys/sshd/ena
/c/sys/access/https/access e
/c/port INT1
        pvid 2
/c/port INT2
        pvid 2
/c/port INT3
        pvid 2
/c/port INT4
        pvid 2
/c/port INT5
        pvid 2
/c/port INT6
        pvid 2
/c/port INT7
        tag ena
        pvid 2
/c/port INT8
        pvid 2
/c/port INT9
        pvid 2
/c/port INT10
        pvid 2
/c/port INT11
        pvid 2
/c/port INT12
        pvid 2
/c/port INT13
        pvid 2
/c/port INT14
        pvid 2
/c/port EXT1
        tag ena
        pvid 2
/c/port EXT2
        tag ena
        pvid 2
/c/port EXT4
```

```
        tag ena
/c/port EXT6
        tag ena
/c/l2/vlan 1
        def EXT1 EXT2 EXT3 EXT4 EXT5 EXT6
/c/l2/vlan 2
        ena
        name "VLAN 2"
        def INT1 INT2 INT3 INT4 INT5 INT6 INT7 INT8 INT9 INT10 INT11 INT12 IN
 INT14 EXT1 EXT2 EXT4 EXT6
/c/l2/vlan 3
        ena
        name "VLAN 3"
        def INT7 EXT1 EXT2 EXT4 EXT6
/c/l2/mrst/on
/c/l2/mrst/mode mstp
/c/l2/mrst/rev 1/maxhop 20
/c/l2/mrst/name "TENGIG"
/c/l2/mrst/cist/add 1
/c/l2/mrst/cist/port EXT1/cost 1000
/c/l2/mrst/cist/port EXT2/cost 1000
/c/l2/stg 1/port EXT1/cost 1000
/c/l2/stg 2/clear
/c/l2/stg 2/add 2
/c/l2/stg 2/port EXT1/cost 1000
/c/l2/stg 2/port EXT2/cost 1000
/c/l2/stg 3/clear
/c/l2/stg 3/add 3
/c/l2/stg 3/port EXT1/cost 1000
/c/l2/stg 3/port EXT2/cost 1000
/c/l2/trunk 1
        ena
        add EXT1
        add EXT2
/c/l2/lacp/port EXT5
        mode active
        adminkey 8
/c/l2/lacp/port EXT6
        mode active
        adminkey 99
/c/l3/if 1
        ena
        addr 10.10.10.1
        mask 255.255.255.0
        broad 10.10.10.255
        vlan 2
/c/l3/if 2
        ena
        addr 10.10.3.1
        mask 255.255.255.0
        broad 10.10.3.255
        vlan 3
/c/l3/if 249
        ena
        addr 9.42.171.246
```

```
        mask 255.255.255.0
        broad 9.42.171.255
/c/l3/gw 1
        ena
        addr 10.10.10.200
/c/l3/gw 253
        ena
        addr 9.42.170.1
/c/pmirr/monport INT6
        add EXT1 both
/c/pmirr/monport INT6
        add EXT2 both
/
script end  /**** DO NOT EDIT THIS LINE!
```

Example: A-2 C6509

```
Current configuration : 3313 bytes
!
! NVRAM config last updated at 08:43:38 EST Wed May 16 2007
!
version 12.2
service timestamps debug datetime localtime
service timestamps log datetime localtime
no service password-encryption
service counters max age 10
!
hostname c6509_A
!
boot system flash sup-bootflash:s72033-pk9sv-mz.122-18.SXD7b.bin
boot system flash sup-bootflash:s72033-pk9sv-mz.122-17d.SXB8.bin
logging snmp-authfail
enable secret 5 $1$IHBu$FcKhr9.8b9OJeoBZUvfoU1
!
no aaa new-model
clock timezone EST -4
ip subnet-zero
!
!
no ip domain-lookup
!
mls ip multicast flow-stat-timer 9
no mls flow ip
no mls flow ipv6
mls cef error action freeze
!
!
!
!
!
spanning-tree mode mst
no spanning-tree optimize bpdu transmission
no spanning-tree vlan 1
spanning-tree vlan 1-3 priority 8192
```

```
!
spanning-tree mst configuration
 name TENGIG
 revision 1
 instance 2 vlan 2
 instance 3 vlan 3
 instance 10 vlan 4-4094
!
spanning-tree mst 0-3 priority 24576
diagnostic cns publish cisco.cns.device.diag_results
diagnostic cns subscribe cisco.cns.device.diag_commands
!
redundancy
 mode sso
 main-cpu
  auto-sync running-config
  auto-sync standard
!
vlan internal allocation policy ascending
vlan access-log ratelimit 2000
!
!
interface Port-channel1
 no ip address
 switchport
 switchport access vlan 2
 switchport trunk encapsulation dot1q
 switchport trunk native vlan 2
 switchport trunk allowed vlan 1-3
 switchport mode trunk
 switchport nonegotiate
!
interface Port-channel2
 no ip address
 switchport
 switchport access vlan 2
 switchport trunk encapsulation dot1q
 switchport trunk native vlan 2
 switchport mode trunk
 switchport nonegotiate
!
interface TenGigabitEthernet3/1
 no ip address
 switchport
 switchport access vlan 2
 switchport trunk encapsulation dot1q
 switchport trunk native vlan 2
 switchport trunk allowed vlan 1-3
 switchport mode trunk
 switchport nonegotiate
 channel-group 1 mode on
!
interface TenGigabitEthernet3/2
 no ip address
 switchport
```

```
 switchport access vlan 2
 switchport trunk encapsulation dot1q
 switchport trunk native vlan 2
 switchport trunk allowed vlan 1-3
 switchport mode trunk
 switchport nonegotiate
 channel-group 1 mode on
!
interface TenGigabitEthernet3/3
 no ip address
 switchport
 switchport access vlan 2
 switchport trunk encapsulation dot1q
 switchport trunk native vlan 2
 switchport mode trunk
 switchport nonegotiate
 channel-protocol lacp
 channel-group 2 mode active
!
interface TenGigabitEthernet3/4
 no ip address
 switchport
 switchport access vlan 2
 switchport trunk encapsulation dot1q
 switchport trunk native vlan 2
 switchport mode trunk
 switchport nonegotiate
 channel-protocol lacp
 channel-group 2 mode active
!
interface GigabitEthernet5/1
 no ip address
 switchport
 switchport access vlan 3
 switchport trunk encapsulation dot1q
 switchport trunk native vlan 3
 switchport mode trunk
 switchport nonegotiate
!
interface GigabitEthernet5/2
 no ip address
 media-type rj45
 switchport
 switchport mode access
 spanning-tree bpdufilter enable
!
interface Vlan1
 ip address 9.42.170.243 255.255.254.0
!
interface Vlan2
 ip address 10.10.10.200 255.255.255.0
!
interface Vlan3
 ip address 10.10.3.200 255.255.255.0
!
```

```
ip classless
ip route 0.0.0.0 0.0.0.0 9.42.170.1
no ip http server
!
!
!
!
!
!
control-plane
!
!
!
line con 0
line vty 0 4
 password itsOral
 login
!
end
```

Example: A-3 10GESM2

```
>> Main# /c/d
script start "Nortel 10Gb Ethernet Switch Module for IBM BladeCenter" 5  /**** D
O NOT EDIT THIS LINE!
/* Configuration dump taken  9:48:28 Wed May 16, 2007
/* Version 1.0.1.1,  Base MAC address 00:17:ef:c1:6f:00
/c/port INT1
        pvid 2
/c/port INT2
        pvid 2
/c/port INT3
        pvid 2
/c/port INT4
        pvid 2
/c/port INT5
        pvid 2
/c/port INT6
        pvid 2
/c/port INT7
        tag ena
        pvid 2
/c/port INT8
        pvid 2
/c/port INT9
        pvid 2
/c/port INT10
        pvid 2
/c/port INT11
        pvid 2
/c/port INT12
        pvid 2
/c/port INT13
        pvid 2
```

```
/c/port INT14
        pvid 2
/c/port EXT1
        tag ena
        pvid 2
/c/port EXT2
        tag ena
        pvid 2
/c/port EXT4
        tag ena
/c/port EXT6
        tag ena
/c/l2/vlan 1
        def EXT1 EXT2 EXT3 EXT4 EXT5 EXT6
/c/l2/vlan 2
        ena
        name "VLAN 2"
        def INT1 INT2 INT3 INT4 INT5 INT6 INT7 INT8 INT9 INT10 INT11 INT12 INT13
 INT14 EXT1 EXT2 EXT4 EXT6
/c/l2/vlan 3
        ena
        name "VLAN 3"
        def INT7 EXT1 EXT2 EXT4 EXT6
/c/l2/mrst/on
/c/l2/mrst/mode mstp
/c/l2/mrst/rev 1/maxhop 20
/c/l2/mrst/name "TENGIG"
/c/l2/mrst/cist/add 1
/c/l2/mrst/cist/port EXT1/prior 64 /cost 1000
/c/l2/mrst/cist/port EXT2/prior 64 /cost 1000
/c/l2/stg 2/clear
/c/l2/stg 2/add 2
/c/l2/stg 2/port EXT1/prior 64 /cost 1000
/c/l2/stg 2/port EXT2/prior 64 /cost 1000
/c/l2/stg 3/clear
/c/l2/stg 3/add 3
/c/l2/stg 3/port EXT1/cost 1000
/c/l2/stg 3/port EXT2/cost 1000
/c/l2/trunk 1
        dis
        add EXT1
        add EXT2
/c/l2/lacp/port EXT1
        mode active
        adminkey 10
/c/l2/lacp/port EXT2
        mode active
        adminkey 10
/c/l3/if 1
        ena
        addr 10.10.10.2
        mask 255.255.255.0
        broad 10.10.10.255
        vlan 2
/c/l3/if 2
```

```
        ena
        addr 10.10.3.2
        mask 255.255.255.0
        broad 10.10.3.255
        vlan 3
/c/l3/gw 1
        ena
        addr 10.10.10.200
/
script end  /**** DO NOT EDIT THIS LINE!
```

Example: A-4 C6506

```
Current configuration : 2741 bytes
!
! NVRAM config last updated at 08:45:03 EDT Wed May 16 2007
!
version 12.2
service timestamps debug datetime localtime
service timestamps log datetime localtime
no service password-encryption
service counters max age 10
!
hostname c6506
!
boot system flash
logging snmp-authfail
enable secret 5 $1$RZs1$2ck62Q8glJ1dk/jRYQaGr1
!
no aaa new-model
clock timezone EDT -4
ip subnet-zero
!
!
!
mls ip multicast flow-stat-timer 9
no mls flow ip
no mls flow ipv6
mls cef error action freeze
!
!
!
!
!
spanning-tree mode mst
no spanning-tree optimize bpdu transmission
!
spanning-tree mst configuration
 name TENGIG
 revision 1
 instance 2 vlan 2
 instance 3 vlan 3
 instance 10 vlan 4-4094
!
```

```
diagnostic cns publish cisco.cns.device.diag_results
diagnostic cns subscribe cisco.cns.device.diag_commands
!
redundancy
 mode sso
 main-cpu
  auto-sync running-config
  auto-sync standard
!
vlan internal allocation policy ascending
vlan access-log ratelimit 2000
!
!
interface Port-channel1
 no ip address
 switchport
 switchport trunk encapsulation dot1q
 switchport trunk native vlan 2
 switchport trunk allowed vlan 1-3
 switchport mode trunk
 switchport nonegotiate
!
interface TenGigabitEthernet2/1
 no ip address
 switchport
 switchport trunk encapsulation dot1q
 switchport trunk allowed vlan 1-3
 switchport mode trunk
 switchport nonegotiate
!
interface TenGigabitEthernet2/2
 no ip address
!
interface TenGigabitEthernet2/3
 no ip address
 shutdown
!
interface TenGigabitEthernet2/4
 no ip address
 shutdown
!
interface TenGigabitEthernet3/1
 no ip address
 switchport
 switchport trunk encapsulation dot1q
 switchport trunk native vlan 2
 switchport trunk allowed vlan 1-3
 switchport mode trunk
 switchport nonegotiate
 channel-group 1 mode active
!
interface TenGigabitEthernet3/2
 no ip address
 switchport
 switchport trunk encapsulation dot1q
```

```
 switchport trunk native vlan 2
 switchport trunk allowed vlan 1-3
 switchport mode trunk
 switchport nonegotiate
 channel-group 1 mode active
!
interface TenGigabitEthernet3/3
 no ip address
!
interface TenGigabitEthernet3/4
 no ip address
!
interface GigabitEthernet6/1
 no ip address
 switchport
 switchport access vlan 3
 switchport trunk encapsulation dot1q
 switchport trunk native vlan 3
 switchport trunk allowed vlan 1-3
 switchport mode trunk
 switchport nonegotiate
!
interface GigabitEthernet6/2
 no ip address
 media-type rj45
 switchport
 switchport access vlan 3
!
interface Vlan1
 no ip address
 shutdown
!
interface Vlan2
 ip address 10.10.10.201 255.255.255.0
!
interface Vlan3
 ip address 10.10.3.201 255.255.255.0
!
ip classless
ip route 0.0.0.0 0.0.0.0 10.10.3.200
no ip http server
!
!
!
!
!
!
control-plane
!
!
!
line con 0
line vty 0 4
 password itsOral
 login
```

```
!
end

c6506#
```

A.1.2 Configurations for 6.3.2, “Layer 2 configuration with 802.1D, Cisco PVST+ and Nortel extensions” on page 102

The following examples show switch configuration dumps from the topology using Port Fast with MSTP.

Example: A-5 10GESM1

```
>> Main# /c/d
script start "Nortel 10Gb Ethernet Switch Module for IBM BladeCenter" 5  /**** DO
NOT EDIT THIS LINE!
/* Configuration dump taken  4:34:25 Sun Jan  1, 2000
/* Version 1.0.1.1,  Base MAC address 00:17:ef:71:e0:00
/c/sys/sshd/ena
/c/sys/access/https/access e
/c/port INT1
        pvid 10
/c/port INT2
        pvid 10
/c/port INT3
        pvid 10
/c/port INT4
        pvid 10
/c/port INT5
        pvid 10
/c/port INT6
        pvid 10
/c/port INT7
        tag ena
        pvid 10
/c/port INT8
        pvid 10
/c/port INT9
        pvid 10
/c/port INT10
        pvid 10
/c/port INT11
        pvid 10
/c/port INT12
        pvid 10
/c/port INT13
        pvid 10
/c/port INT14
        pvid 10
/c/port EXT1
        tag ena
        pvid 10
        fastfwd ena
```

```
/c/port EXT2
        tag ena
        pvid 10
        fastfwd ena
/c/port EXT3
        pvid 10
/c/port EXT4
        tag ena
/c/port EXT5
        pvid 10
/c/port EXT6
        tag ena
/c/l2/vlan 1
        def EXT1 EXT2 EXT4 EXT6
/c/l2/vlan 3
        ena
        name "VLAN 3"
        def INT7 EXT1 EXT2 EXT4 EXT6
/c/l2/vlan 10
        ena
        name "VLAN 10"
        def INT1 INT2 INT3 INT4 INT5 INT6 INT7 INT8 INT9 INT10 INT11 INT12 INT13
INT14 EXT1 EXT2 EXT3 EXT4 EXT5 EXT6
/c/l2/mrst/mode mstp
/c/l2/mrst/rev 1/maxhop 20
/c/l2/mrst/name "TENGIG"
/c/l2/mrst/cist/port EXT1/cost 1000
/c/l2/mrst/cist/port EXT2/cost 1000
/c/l2/stg 2/off
/c/l2/stg 3/clear
/c/l2/stg 3/add 3
/c/l2/stg 10/clear
/c/l2/stg 10/add 10
/c/l2/lacp/port EXT5
        adminkey 8
/c/l2/lacp/port EXT6
        adminkey 99
/c/l3/if 1
        ena
        addr 10.10.10.1
        mask 255.255.255.0
        broad 10.10.10.255
        vlan 10
/c/l3/if 2
        ena
        addr 10.10.3.1
        mask 255.255.255.0
        broad 10.10.3.255
        vlan 3
/c/l3/if 249
        ena
        addr 9.42.171.246
        mask 255.255.255.0
        broad 9.42.171.255
/c/l3/gw 253
```

```
        ena
        addr 9.42.170.1
/c/pmirr/monport INT6
        add EXT1 both
/c/pmirr/monport INT6
        add EXT2 both
/
script end  /**** DO NOT EDIT THIS LINE!

>> Configuration#
```

Example: A-6 C6509

```
c6509_A#sh run
Building configuration...

Current configuration : 2475 bytes
!
! Last configuration change at 13:57:25 EST Thu May 17 2007
! NVRAM config last updated at 13:50:44 EST Thu May 17 2007
!
version 12.2
service timestamps debug datetime localtime
service timestamps log datetime localtime
no service password-encryption
service counters max age 10
!
hostname c6509_A
!
boot system flash sup-bootflash:s72033-pk9sv-mz.122-18.SXD7b.bin
boot system flash sup-bootflash:s72033-pk9sv-mz.122-17d.SXB8.bin
logging snmp-authfail
enable secret 5 $1$IHBu$FcKhr9.8b9OJeoBZUvfoU1
!
no aaa new-model
clock timezone EST -4
ip subnet-zero
!
!
no ip domain-lookup
!
mls ip multicast flow-stat-timer 9
no mls flow ip
no mls flow ipv6
mls cef error action freeze
!
!
!
!
!
spanning-tree mode mst
no spanning-tree optimize bpdu transmission
spanning-tree vlan 1-3 priority 8192
```

```
!
spanning-tree mst configuration
 name TENGIG
 revision 1
 instance 3 vlan 3
 instance 10 vlan 10
!
spanning-tree mst 0-3 priority 24576
diagnostic cns publish cisco.cns.device.diag_results
diagnostic cns subscribe cisco.cns.device.diag_commands
!
redundancy
 mode sso
 main-cpu
  auto-sync running-config
  auto-sync standard
!
vlan internal allocation policy ascending
vlan access-log ratelimit 2000
!
!
interface TenGigabitEthernet3/1
 no ip address
 switchport
 switchport access vlan 10
 switchport trunk encapsulation dot1q
 switchport trunk native vlan 10
 switchport mode trunk
 switchport nonegotiate
!
interface TenGigabitEthernet3/2
 no ip address
 switchport
 switchport access vlan 10
 switchport trunk encapsulation dot1q
 switchport trunk native vlan 10
 switchport mode trunk
 switchport nonegotiate
!
interface TenGigabitEthernet3/3
 no ip address
 shutdown
!
interface TenGigabitEthernet3/4
 no ip address
 shutdown
!
interface GigabitEthernet5/1
 no ip address
 switchport
 switchport access vlan 3
 switchport trunk encapsulation dot1q
 switchport trunk native vlan 3
 switchport mode trunk
 switchport nonegotiate
```

```
 spanning-tree cost 10
!
interface GigabitEthernet5/2
 no ip address
 media-type rj45
 switchport
 switchport mode access
 spanning-tree bpdufilter enable
!
interface Vlan1
 ip address 9.42.170.243 255.255.254.0
!
interface Vlan3
 ip address 10.10.3.200 255.255.255.0
!
interface Vlan10
 ip address 10.10.10.200 255.255.255.0
!
router rip
 version 2
 network 10.0.0.0
 network 172.16.0.0
!
ip classless
ip route 0.0.0.0 0.0.0.0 9.42.170.1
no ip http server
!
!
!
!
!
!
control-plane
!
!
!
line con 0
line vty 0 4
 password itsOral
 login
!
end

c6509_A#
```

Example: A-7 10GESM2

```
>> Main# /c/d
script start "Nortel 10Gb Ethernet Switch Module for IBM BladeCenter" 5  /**** DO
NOT EDIT THIS LINE!
/* Configuration dump taken 23:35:26 Sat Dec 31, 2000
/* Version 1.0.1.1,  Base MAC address 00:17:ef:c1:6f:00
/c/port INT1
```

```
        pvid 10
/c/port INT2
        pvid 10
/c/port INT3
        pvid 10
/c/port INT4
        pvid 10
/c/port INT5
        pvid 10
/c/port INT6
        pvid 10
/c/port INT7
        tag ena
        pvid 10
/c/port INT8
        pvid 10
/c/port INT9
        pvid 10
/c/port INT10
        pvid 10
/c/port INT11
        pvid 10
/c/port INT12
        pvid 10
/c/port INT13
        pvid 10
/c/port INT14
        pvid 10
/c/port EXT1
        tag ena
        pvid 10
        fastfwd ena
/c/port EXT2
        tag ena
        pvid 10
        fastfwd ena
/c/port EXT3
        pvid 10
/c/port EXT4
        tag ena
/c/port EXT5
        pvid 10
/c/port EXT6
        tag ena
/c/l2/vlan 1
        def EXT1 EXT2 EXT4 EXT6
/c/l2/vlan 3
        ena
        name "VLAN 3"
        def INT7 EXT1 EXT2 EXT4 EXT6
/c/l2/vlan 10
        ena
        name "VLAN 10"
        def INT1 INT2 INT3 INT4 INT5 INT6 INT7 INT8 INT9 INT10 INT11 INT12 INT13
INT14 EXT1 EXT2 EXT3 EXT4 EXT5 EXT6
```

```
/c/l2/mrst/mode mstp
/c/l2/mrst/rev 1/maxhop 20
/c/l2/mrst/name "TENGIG"
/c/l2/mrst/cist/port EXT1/prior 64 /cost 1000
/c/l2/mrst/cist/port EXT2/prior 64 /cost 1000
/c/l2/stg 2/port EXT1/prior 64
/c/l2/stg 2/port EXT2/prior 64
/c/l2/stg 3/clear
/c/l2/stg 3/add 3
/c/l2/stg 10/clear
/c/l2/stg 10/add 10
/c/l2/lacp/port EXT1
        adminkey 10
/c/l2/lacp/port EXT2
        adminkey 10
/c/l3/if 1
        ena
        addr 10.10.10.2
        mask 255.255.255.0
        broad 10.10.10.255
        vlan 10
/c/l3/if 2
        ena
        addr 10.10.3.2
        mask 255.255.255.0
        broad 10.10.3.255
        vlan 3
/
script end  /**** DO NOT EDIT THIS LINE!

>> Configuration#
```

Example: A-8 C6506

```
c6506#sh run
Building configuration...

Current configuration : 2615 bytes
!
! Last configuration change at 13:55:15 EDT Thu May 17 2007
! NVRAM config last updated at 13:50:40 EDT Thu May 17 2007
!
version 12.2
service timestamps debug datetime localtime
service timestamps log datetime localtime
no service password-encryption
service counters max age 10
!
hostname c6506
!
boot system flash
logging snmp-authfail
enable secret 5 $1$RZs1$2ck62Q8glJ1dk/jRYQaGr1
```

```
!
no aaa new-model
clock timezone EDT -4
ip subnet-zero
!
!
!
mls ip multicast flow-stat-timer 9
no mls flow ip
no mls flow ipv6
mls cef error action freeze
!
!
!
!
!
spanning-tree mode mst
no spanning-tree optimize bpdu transmission
!
spanning-tree mst configuration
 name TENGIG
 revision 1
 instance 3 vlan 3
 instance 10 vlan 10
!
diagnostic cns publish cisco.cns.device.diag_results
diagnostic cns subscribe cisco.cns.device.diag_commands
!
redundancy
 mode sso
 main-cpu
  auto-sync running-config
  auto-sync standard
!
vlan internal allocation policy ascending
vlan access-log ratelimit 2000
!
!
interface Loopback192
 ip address 192.168.1.1 255.255.255.0
!
interface TenGigabitEthernet2/1
 no ip address
 switchport
 switchport trunk encapsulation dot1q
 switchport trunk allowed vlan 1-3
 switchport mode trunk
 switchport nonegotiate
!
interface TenGigabitEthernet2/2
 no ip address
!
interface TenGigabitEthernet2/3
 no ip address
 shutdown
```

```
!
interface TenGigabitEthernet2/4
 no ip address
 shutdown
!
interface TenGigabitEthernet3/1
 no ip address
 switchport
 switchport access vlan 10
 switchport trunk encapsulation dot1q
 switchport trunk native vlan 10
 switchport mode trunk
 switchport nonegotiate
!
interface TenGigabitEthernet3/2
 no ip address
 switchport
 switchport access vlan 10
 switchport trunk encapsulation dot1q
 switchport trunk native vlan 10
 switchport mode trunk
 switchport nonegotiate
!
interface TenGigabitEthernet3/3
 no ip address
!
interface TenGigabitEthernet3/4
 no ip address
!
interface GigabitEthernet6/1
 no ip address
 switchport
 switchport access vlan 3
 switchport trunk encapsulation dot1q
 switchport trunk native vlan 3
 switchport mode trunk
 switchport nonegotiate
 spanning-tree cost 10
!
interface GigabitEthernet6/2
 no ip address
 media-type rj45
 switchport
 switchport access vlan 3
!
interface Vlan1
 no ip address
 shutdown
!
interface Vlan3
 ip address 10.10.3.201 255.255.255.0
!
interface Vlan10
 ip address 10.10.10.201 255.255.255.0
!
```

```
router rip
 version 2
 network 10.0.0.0
 network 192.168.1.0
!
ip classless
ip route 0.0.0.0 0.0.0.0 10.10.3.200
no ip http server
!
!
!
!
!
!
control-plane
!
!
!
line con 0
line vty 0 4
 password itsOral
 login
!
end

c6506#
```

A.1.3 Configurations for 6.4.6, “L3 configuration with VRRP and static routes” on page 113

The following examples show switch configuration dumps from the full mesh topology using VRRP and static routes.

Example: A-9 10GESM1

```
>> Main# /c/d
script start "Nortel 10Gb Ethernet Switch Module for IBM BladeCenter" 5  /**** DO
NOT EDIT THIS LINE!
/* Configuration dump taken  4:33:34 Mon Jan  2, 2000
/* Version 1.0.1.1,  Base MAC address 00:17:ef:71:e0:00
/c/sys/sshd/ena
/c/sys/access/https/access e
/c/port INT1
        pvid 10
/c/port INT2
        pvid 10
/c/port INT3
        pvid 10
/c/port INT4
        pvid 10
/c/port INT5
        pvid 10
```

```
/c/port INT6
        pvid 10
/c/port INT7
        tag ena
        pvid 10
/c/port INT8
        pvid 10
/c/port INT9
        pvid 10
/c/port INT10
        pvid 10
/c/port INT11
        pvid 10
/c/port INT12
        pvid 10
/c/port INT13
        pvid 10
/c/port INT14
        pvid 10
/c/port EXT1
        pvid 5
        fastfwd ena
/c/port EXT2
        pvid 15
        fastfwd ena
/c/port EXT3
        pvid 10
/c/port EXT4
        tag ena
        pvid 10
/c/port EXT5
        pvid 10
/c/port EXT6
        tag ena
        pvid 10
/c/l2/vlan 1
        def 0
/c/l2/vlan 3
        ena
        name "VLAN 3"
        def INT7 EXT4 EXT6
/c/l2/vlan 5
        ena
        name "VLAN 5"
        def EXT1
/c/l2/vlan 10
        ena
        name "VLAN 10"
        def INT1 INT2 INT3 INT4 INT5 INT6 INT7 INT8 INT9 INT10 INT11 INT12 INT13
INT14 EXT3 EXT4 EXT5 EXT6
/c/l2/vlan 15
        ena
        name "VLAN 15"
        def EXT2
/c/l2/mrst/mode mstp
```

```
/c/l2/mrst/rev 1/maxhop 20
/c/l2/mrst/name "TENGIG"
/c/l2/mrst/cist/port EXT1/cost 1000
/c/l2/mrst/cist/port EXT2/cost 1000
/c/l2/stg 1/off
/c/l2/stg 1/clear
/c/l2/stg 1/add 1 5 15
/c/l2/stg 2/off
/c/l2/stg 3/off
/c/l2/stg 3/clear
/c/l2/stg 3/add 3
/c/l2/stg 10/off
/c/l2/stg 10/clear
/c/l2/stg 10/add 10
/c/l2/trunk 1
        ena
        add EXT1
/c/l2/trunk 2
        ena
        add EXT2
/c/l2/lacp/port EXT5
        adminkey 8
/c/l2/lacp/port EXT6
        adminkey 99
/c/l2/failovr/on
/c/l2/failovr/trigger 1
        ena
        limit 0
/c/l2/failovr/trigger 1/amon
        addtrnk 1
        addtrnk 2
/c/l3/if 1
        ena
        addr 10.10.10.1
        mask 255.255.255.0
        broad 10.10.10.255
        vlan 10
/c/l3/if 2
        ena
        addr 10.10.3.1
        mask 255.255.255.0
        broad 10.10.3.255
        vlan 3
/c/l3/if 5
        ena
        addr 10.10.5.1
        mask 255.255.255.0
        broad 10.10.5.255
        vlan 5
/c/l3/if 15
        ena
        addr 10.10.15.1
        mask 255.255.255.0
        broad 10.10.15.255
        vlan 15
```

```
/c/l3/if 249
        ena
        addr 9.42.171.246
        mask 255.255.255.0
        broad 9.42.171.255
/c/l3/gw 1
        ena
        addr 10.10.15.201
/c/l3/gw 2
        ena
        addr 10.10.5.200
/c/l3/gw 253
        ena
        addr 9.42.170.1
/c/l3/vrrp/on
/c/l3/vrrp/vr 3
        ena
        vrid 3
        if 2
        prio 103
        addr 10.10.3.254
/c/l3/vrrp/vr 3/track
                ports e
/c/l3/vrrp/vr 10
        ena
        vrid 10
        if 1
        prio 103
        addr 10.10.10.254
/c/l3/vrrp/vr 10/track
                ports e
/
script end  /**** DO NOT EDIT THIS LINE!

>> Configuration#
```

Example: A-10 C6509

```
c6509_A#sh run
Building configuration...

Current configuration : 2145 bytes
!
version 12.2
service timestamps debug datetime localtime
service timestamps log datetime localtime
no service password-encryption
service counters max age 10
!
hostname c6509_A
!
boot system flash sup-bootflash:s72033-pk9sv-mz.122-18.SXD7b.bin
boot system flash sup-bootflash:s72033-pk9sv-mz.122-17d.SXB8.bin
```

```
logging snmp-authfail
enable secret 5 $1$IHBu$FcKhr9.8b9OJeoBZUvfoU1
!
no aaa new-model
clock timezone EST -4
ip subnet-zero
!
!
no ip domain-lookup
!
mls ip multicast flow-stat-timer 9
no mls flow ip
no mls flow ipv6
mls cef error action freeze
!
!
!
!
!
spanning-tree mode pvst
no spanning-tree optimize bpdu transmission
spanning-tree vlan 1-3 priority 8192
!
spanning-tree mst configuration
 name TENGIG
 revision 1
 instance 3 vlan 3
 instance 10 vlan 10
!
spanning-tree mst 0-3 priority 24576
diagnostic cns publish cisco.cns.device.diag_results
diagnostic cns subscribe cisco.cns.device.diag_commands
!
redundancy
 mode sso
 main-cpu
  auto-sync running-config
  auto-sync standard
!
vlan internal allocation policy ascending
vlan access-log ratelimit 2000
!
!
interface TenGigabitEthernet3/1
 ip address 10.10.5.200 255.255.255.0
!
interface TenGigabitEthernet3/2
 ip address 10.10.25.200 255.255.255.0
!
interface TenGigabitEthernet3/3
 no ip address
 shutdown
!
interface TenGigabitEthernet3/4
 no ip address
```

```
 shutdown
!
interface GigabitEthernet5/1
 no ip address
 switchport
 switchport access vlan 3
 switchport trunk encapsulation dot1q
 switchport trunk native vlan 3
 switchport mode trunk
 switchport nonegotiate
 spanning-tree cost 10
!
interface GigabitEthernet5/2
 no ip address
 media-type rj45
 switchport
 switchport mode access
 spanning-tree bpdufilter enable
!
interface Vlan1
 ip address 9.42.170.243 255.255.254.0
!
interface Vlan30
 ip address 10.10.30.200 255.255.255.0
!
ip classless
ip route 0.0.0.0 0.0.0.0 9.42.170.1
ip route 10.10.3.0 255.255.255.0 10.10.25.2
ip route 10.10.3.0 255.255.255.0 10.10.5.1
ip route 10.10.10.0 255.255.255.0 10.10.5.1
ip route 10.10.10.0 255.255.255.0 10.10.25.2
no ip http server
!
!
!
!
!
!
control-plane
!
!
!
line con 0
line vty 0 4
 password itsOral
 login
!
end

c6509_A#
```

Example: A-11 10GESM2

```
>> Main# /c/d
script start "Nortel 10Gb Ethernet Switch Module for IBM BladeCenter" 5  /**** DO
NOT EDIT THIS LINE!
/* Configuration dump taken 23:34:34 Sun Jan  1, 2000
/* Version 1.0.1.1,  Base MAC address 00:17:ef:c1:6f:00
/c/port INT1
        pvid 10
/c/port INT2
        pvid 10
/c/port INT3
        pvid 10
/c/port INT4
        pvid 10
/c/port INT5
        pvid 10
/c/port INT6
        pvid 10
/c/port INT7
        tag ena
        pvid 10
/c/port INT8
        pvid 10
/c/port INT9
        pvid 10
/c/port INT10
        pvid 10
/c/port INT11
        pvid 10
/c/port INT12
        pvid 10
/c/port INT13
        pvid 10
/c/port INT14
        pvid 10
/c/port EXT1
        pvid 20
        fastfwd ena
/c/port EXT2
        pvid 25
        fastfwd ena
/c/port EXT3
        pvid 10
/c/port EXT4
        tag ena
        pvid 10
/c/port EXT5
        pvid 10
/c/port EXT6
        tag ena
        pvid 10
/c/l2/vlan 1
        def 0
/c/l2/vlan 3
        ena
```

```
        name "VLAN 3"
        def INT7 EXT4 EXT6
/c/l2/vlan 10
        ena
        name "VLAN 10"
        def INT1 INT2 INT3 INT4 INT5 INT6 INT7 INT8 INT9 INT10 INT11 INT12 INT13
INT14 EXT3 EXT4 EXT5 EXT6
/c/l2/vlan 20
        ena
        name "VLAN 20"
        def EXT1
/c/l2/vlan 25
        ena
        name "VLAN 25"
        def EXT2
/c/l2/mrst/mode mstp
/c/l2/mrst/rev 1/maxhop 20
/c/l2/mrst/name "TENGIG"
/c/l2/mrst/cist/port EXT1/prior 64 /cost 1000
/c/l2/mrst/cist/port EXT2/prior 64 /cost 1000
/c/l2/stg 1/off
/c/l2/stg 1/clear
/c/l2/stg 1/add 1 20 25
/c/l2/stg 2/port EXT1/prior 64
/c/l2/stg 2/port EXT2/prior 64
/c/l2/stg 3/off
/c/l2/stg 3/clear
/c/l2/stg 3/add 3
/c/l2/stg 10/off
/c/l2/stg 10/clear
/c/l2/stg 10/add 10
/c/l2/trunk 1
        ena
        add EXT1
/c/l2/trunk 2
        ena
        add EXT2
/c/l2/lacp/port EXT1
        adminkey 10
/c/l2/lacp/port EXT2
        adminkey 10
/c/l2/failovr/on
/c/l2/failovr/trigger 1
        ena
        limit 0
/c/l2/failovr/trigger 1/amon
        addtrnk 1
        addtrnk 2
/c/l3/if 1
        ena
        addr 10.10.10.2
        mask 255.255.255.0
        broad 10.10.10.255
        vlan 10
/c/l3/if 2
```

```
        ena
        addr 10.10.3.2
        mask 255.255.255.0
        broad 10.10.3.255
        vlan 3
/c/l3/if 20
        ena
        addr 10.10.20.2
        mask 255.255.255.0
        broad 10.10.20.255
        vlan 20
/c/l3/if 25
        ena
        addr 10.10.25.2
        mask 255.255.255.0
        broad 10.10.25.255
        vlan 25
/c/l3/gw 1
        ena
        addr 10.10.25.200
/c/l3/gw 2
        ena
        addr 10.10.20.201
/c/l3/vrrp/on
/c/l3/vrrp/vr 3
        ena
        vrid 3
        if 2
        addr 10.10.3.254
/c/l3/vrrp/vr 3/track
                ports e
/c/l3/vrrp/vr 10
        ena
        vrid 10
        if 1
        addr 10.10.10.254
/c/l3/vrrp/vr 10/track
                ports e
/
script end  /**** DO NOT EDIT THIS LINE!

>> Configuration#
```

Example: A-12 C6506

```
c6506#sh run
Building configuration...

Current configuration : 2802 bytes
!
version 12.2
service timestamps debug datetime localtime
service timestamps log datetime localtime
```

```
no service password-encryption
service counters max age 10
!
hostname c6506
!
boot system flash
logging snmp-authfail
enable secret 5 $1$RZs1$2ck62Q8glJ1dk/jRYQaGr1
!
no aaa new-model
clock timezone EDT -4
ip subnet-zero
!
!
!
mls ip multicast flow-stat-timer 9
no mls flow ip
no mls flow ipv6
mls cef error action freeze
!
!
!
!
!
spanning-tree mode pvst
no spanning-tree optimize bpdu transmission
no spanning-tree vlan 15,20
!
spanning-tree mst configuration
 name TENGIG
 revision 1
 instance 3 vlan 3
 instance 10 vlan 10
!
diagnostic cns publish cisco.cns.device.diag_results
diagnostic cns subscribe cisco.cns.device.diag_commands
!
redundancy
 mode sso
 main-cpu
  auto-sync running-config
  auto-sync standard
!
vlan internal allocation policy ascending
vlan access-log ratelimit 2000
!
!
interface Loopback192
 ip address 192.168.1.1 255.255.255.0
!
interface TenGigabitEthernet2/1
 no ip address
 switchport
 switchport trunk encapsulation dot1q
 switchport trunk allowed vlan 1-3
```

```
 switchport mode trunk
 switchport nonegotiate
!
interface TenGigabitEthernet2/2
 no ip address
!
interface TenGigabitEthernet2/3
 no ip address
 shutdown
!
interface TenGigabitEthernet2/4
 no ip address
 shutdown
!
interface TenGigabitEthernet3/1
 no ip address
 switchport
 switchport access vlan 15
 switchport trunk encapsulation dot1q
 switchport trunk native vlan 10
 switchport mode access
 switchport nonegotiate
!
interface TenGigabitEthernet3/2
 no ip address
 switchport
 switchport access vlan 20
 switchport trunk encapsulation dot1q
 switchport trunk native vlan 10
 switchport mode access
 switchport nonegotiate
!
interface TenGigabitEthernet3/3
 no ip address
!
interface TenGigabitEthernet3/4
 no ip address
!
interface GigabitEthernet6/1
 no ip address
 switchport
 switchport access vlan 3
 switchport trunk encapsulation dot1q
 switchport trunk native vlan 3
 switchport mode trunk
 switchport nonegotiate
 spanning-tree cost 10
!
interface GigabitEthernet6/2
 no ip address
 media-type rj45
 switchport
 switchport access vlan 3
!
interface Vlan1
```

```
 no ip address
 shutdown
!
interface Vlan15
 ip address 10.10.15.201 255.255.255.0
!
interface Vlan20
 ip address 10.10.20.201 255.255.255.0
!
interface Vlan30
 ip address 10.10.30.201 255.255.255.0
!
router rip
 version 2
 network 10.0.0.0
 network 192.168.1.0
!
ip classless
ip route 0.0.0.0 0.0.0.0 10.10.3.200
ip route 9.0.0.0 255.0.0.0 10.10.30.200
ip route 10.10.3.0 255.255.255.0 10.10.15.1
ip route 10.10.3.0 255.255.255.0 10.10.20.2
ip route 10.10.10.0 255.255.255.0 10.10.15.1
ip route 10.10.10.0 255.255.255.0 10.10.20.2
no ip http server
!
!
!
!
!
!
control-plane
!
!
!
line con 0
line vty 0 4
 password itsOral
 login
!
end

c6506#
```

A.1.4 Configurations for 6.4.7, “L3 configuration with VRRP and RIP” on page 119

The following examples show switch configuration dumps from the full mesh topology using VRRP and RIP.

Example: A-13 10GESM1

```
>> Main# /c/d
script start "Nortel 10Gb Ethernet Switch Module for IBM BladeCenter" 5  /**** DO
NOT EDIT THIS LINE!
/* Configuration dump taken 23:22:50 Wed Jan  4, 2000
/* Version 1.0.1.1,  Base MAC address 00:17:ef:71:e0:00
/c/sys/sshd/ena
/c/sys/access/https/access e
/c/port INT1
        pvid 10
/c/port INT2
        pvid 10
/c/port INT3
        pvid 10
/c/port INT4
        pvid 10
/c/port INT5
        pvid 10
/c/port INT6
        pvid 10
/c/port INT7
        tag ena
        pvid 10
/c/port INT8
        pvid 10
/c/port INT9
        pvid 10
/c/port INT10
        pvid 10
/c/port INT11
        pvid 10
/c/port INT12
        pvid 10
/c/port INT13
        pvid 10
/c/port INT14
        pvid 10
/c/port EXT1
        pvid 5
        fastfwd ena
/c/port EXT2
        pvid 15
        fastfwd ena
/c/port EXT3
        pvid 10
/c/port EXT4
        tag ena
        pvid 10
/c/port EXT5
        pvid 10
/c/port EXT6
        tag ena
        pvid 10
/c/l2/vlan 1
        def 0
```

```
/c/l2/vlan 3
        ena
        name "VLAN 3"
        def INT7 EXT4 EXT6
/c/l2/vlan 5
        ena
        name "VLAN 5"
        def EXT1
/c/l2/vlan 10
        ena
        name "VLAN 10"
        def INT1 INT2 INT3 INT4 INT5 INT6 INT7 INT8 INT9 INT10 INT11 INT12 INT13
INT14 EXT3 EXT4 EXT5 EXT6
/c/l2/vlan 15
        ena
        name "VLAN 15"
        def EXT2
/c/l2/mrst/mode mstp
/c/l2/mrst/rev 1/maxhop 20
/c/l2/mrst/name "TENGIG"
/c/l2/mrst/cist/port EXT1/cost 1000
/c/l2/mrst/cist/port EXT2/cost 1000
/c/l2/stg 1/off
/c/l2/stg 1/clear
/c/l2/stg 1/add 1 5 15
/c/l2/stg 2/off
/c/l2/stg 3/off
/c/l2/stg 3/clear
/c/l2/stg 3/add 3
/c/l2/stg 10/off
/c/l2/stg 10/clear
/c/l2/stg 10/add 10
/c/l2/trunk 1
        ena
        add EXT1
/c/l2/trunk 2
        ena
        add EXT2
/c/l2/lacp/port EXT5
        adminkey 8
/c/l2/lacp/port EXT6
        adminkey 99
/c/l2/failovr/on
/c/l2/failovr/trigger 1
        ena
        limit 0
/c/l2/failovr/trigger 1/amon
        addtrnk 1
        addtrnk 2
/c/l3/if 1
        ena
        addr 10.10.10.1
        mask 255.255.255.0
        broad 10.10.10.255
        vlan 10
```

```
/c/l3/if 2
        ena
        addr 10.10.3.1
        mask 255.255.255.0
        broad 10.10.3.255
        vlan 3
/c/l3/if 5
        ena
        addr 10.10.5.1
        mask 255.255.255.0
        broad 10.10.5.255
        vlan 5
/c/l3/if 15
        ena
        addr 10.10.15.1
        mask 255.255.255.0
        broad 10.10.15.255
        vlan 15
/c/l3/if 249
        ena
        addr 9.42.171.246
        mask 255.255.255.0
        broad 9.42.171.255
/c/l3/gw 1
        ena
        addr 10.10.15.201
/c/l3/gw 2
        ena
        addr 10.10.5.200
/c/l3/gw 253
        ena
        addr 9.42.170.1
/c/l3/route
        add 10.0.0.0 255.0.0.0 10.10.15.201 15
/c/l3/vrrp/on
/c/l3/vrrp/vr 3
        ena
        vrid 3
        if 2
        prio 103
        addr 10.10.3.254
/c/l3/vrrp/vr 3/track
                ports e
/c/l3/vrrp/vr 10
        ena
        vrid 10
        if 1
        prio 103
        addr 10.10.10.254
/c/l3/vrrp/vr 10/track
                ports e
/c/l3/rip/updat 30/on
/c/l3/rip/if 1/ena/supply e/listen e/default none/version 2
/c/l3/rip/if 1/poison d/split e/trigg e/metric 1/mcast e
/c/l3/rip/if 1/auth none
```

```
/c/l3/rip/if 2/ena/supply e/listen e/default none/version 2
/c/l3/rip/if 2/poison d/split e/trigg e/metric 1/mcast e
/c/l3/rip/if 2/auth none
/c/l3/rip/if 5/ena/supply e/listen e/default none/version 2
/c/l3/rip/if 5/poison d/split e/trigg e/metric 1/mcast e
/c/l3/rip/if 5/auth none
/c/l3/rip/if 15/ena/supply e/listen e/default none/version 2
/c/l3/rip/if 15/poison d/split e/trigg e/metric 1/mcast e
/c/l3/rip/if 15/auth none
/
script end  /**** DO NOT EDIT THIS LINE!

>> Configuration#
```

Example: A-14 C6509

```
c6509_A#sh run
Building configuration...

Current configuration : 2037 bytes
!
version 12.2
service timestamps debug datetime localtime
service timestamps log datetime localtime
no service password-encryption
service counters max age 10
!
hostname c6509_A
!
boot system flash sup-bootflash:s72033-pk9sv-mz.122-18.SXD7b.bin
boot system flash sup-bootflash:s72033-pk9sv-mz.122-17d.SXB8.bin
logging snmp-authfail
enable secret 5 $1$IHBu$FcKhr9.8b9OJeoBZUvfoU1
!
no aaa new-model
clock timezone EST -4
ip subnet-zero
!
!
no ip domain-lookup
!
mls ip multicast flow-stat-timer 9
no mls flow ip
no mls flow ipv6
mls cef error action freeze
!
!
!
!
!
spanning-tree mode pvst
no spanning-tree optimize bpdu transmission
spanning-tree vlan 1-3 priority 8192
```

```
!
spanning-tree mst configuration
 name TENGIG
 revision 1
 instance 3 vlan 3
 instance 10 vlan 10
!
spanning-tree mst 0-3 priority 24576
diagnostic cns publish cisco.cns.device.diag_results
diagnostic cns subscribe cisco.cns.device.diag_commands
!
redundancy
 mode sso
 main-cpu
  auto-sync running-config
  auto-sync standard
!
vlan internal allocation policy ascending
vlan access-log ratelimit 2000
!
!
interface TenGigabitEthernet3/1
 ip address 10.10.5.200 255.255.255.0
!
interface TenGigabitEthernet3/2
 ip address 10.10.25.200 255.255.255.0
!
interface TenGigabitEthernet3/3
 no ip address
 shutdown
!
interface TenGigabitEthernet3/4
 no ip address
 shutdown
!
interface GigabitEthernet5/1
 no ip address
 switchport
 switchport access vlan 3
 switchport trunk encapsulation dot1q
 switchport trunk native vlan 3
 switchport mode trunk
 switchport nonegotiate
 spanning-tree cost 10
!
interface GigabitEthernet5/2
 no ip address
 media-type rj45
 switchport
 switchport mode access
 spanning-tree bpdufilter enable
!
interface Vlan1
 ip address 9.42.170.243 255.255.254.0
!
```

```
interface Vlan30
 ip address 10.10.30.200 255.255.255.0
!
router rip
 traffic-share min across-interfaces
 network 10.0.0.0
!
ip classless
ip route 0.0.0.0 0.0.0.0 9.42.170.1
no ip http server
!
!
!
!
!
!
control-plane
!
!
!
line con 0
line vty 0 4
 password itsOral
 login
!
end

c6509_A#
```

Example: A-15 10GESM2

```
>> Main# /c/d
script start "Nortel 10Gb Ethernet Switch Module for IBM BladeCenter" 5  /**** DO
NOT EDIT THIS LINE!
/* Configuration dump taken 18:23:40 Wed Jan  4, 2000
/* Version 1.0.1.1,  Base MAC address 00:17:ef:c1:6f:00
/c/port INT1
        pvid 10
/c/port INT2
        pvid 10
/c/port INT3
        pvid 10
/c/port INT4
        pvid 10
/c/port INT5
        pvid 10
/c/port INT6
        pvid 10
/c/port INT7
        tag ena
        pvid 10
/c/port INT8
        pvid 10
```

```
/c/port INT9
        pvid 10
/c/port INT10
        pvid 10
/c/port INT11
        pvid 10
/c/port INT12
        pvid 10
/c/port INT13
        pvid 10
/c/port INT14
        pvid 10
/c/port EXT1
        pvid 20
        fastfwd ena
/c/port EXT2
        pvid 25
        fastfwd ena
/c/port EXT3
        pvid 10
/c/port EXT4
        tag ena
        pvid 10
/c/port EXT5
        pvid 10
/c/port EXT6
        tag ena
        pvid 10
/c/l2/vlan 1
        def 0
/c/l2/vlan 3
        ena
        name "VLAN 3"
        def INT7 EXT4 EXT6
/c/l2/vlan 10
        ena
        name "VLAN 10"
        def INT1 INT2 INT3 INT4 INT5 INT6 INT7 INT8 INT9 INT10 INT11 INT12 INT13
INT14 EXT3 EXT4 EXT5 EXT6
/c/l2/vlan 20
        ena
        name "VLAN 20"
        def EXT1
/c/l2/vlan 25
        ena
        name "VLAN 25"
        def EXT2
/c/l2/mrst/mode mstp
/c/l2/mrst/rev 1/maxhop 20
/c/l2/mrst/name "TENGIG"
/c/l2/mrst/cist/port EXT1/prior 64 /cost 1000
/c/l2/mrst/cist/port EXT2/prior 64 /cost 1000
/c/l2/stg 1/off
/c/l2/stg 1/clear
/c/l2/stg 1/add 1 20 25
```

```
/c/l2/stg 2/port EXT1/prior 64
/c/l2/stg 2/port EXT2/prior 64
/c/l2/stg 3/off
/c/l2/stg 3/clear
/c/l2/stg 3/add 3
/c/l2/stg 10/off
/c/l2/stg 10/clear
/c/l2/stg 10/add 10
/c/l2/trunk 1
        ena
        add EXT1
/c/l2/trunk 2
        ena
        add EXT2
/c/l2/lacp/port EXT1
        adminkey 10
/c/l2/lacp/port EXT2
        adminkey 10
/c/l2/failovr/on
/c/l2/failovr/trigger 1
        ena
        limit 0
/c/l2/failovr/trigger 1/amon
        addtrnk 1
        addtrnk 2
/c/l3/if 1
        ena
        addr 10.10.10.2
        mask 255.255.255.0
        broad 10.10.10.255
        vlan 10
/c/l3/if 2
        ena
        addr 10.10.3.2
        mask 255.255.255.0
        broad 10.10.3.255
        vlan 3
/c/l3/if 20
        ena
        addr 10.10.20.2
        mask 255.255.255.0
        broad 10.10.20.255
        vlan 20
/c/l3/if 25
        ena
        addr 10.10.25.2
        mask 255.255.255.0
        broad 10.10.25.255
        vlan 25
/c/l3/gw 1
        ena
        addr 10.10.25.200
/c/l3/gw 2
        ena
        addr 10.10.20.201
```

```
/c/l3/vrrp/on
/c/l3/vrrp/vr 3
        ena
        vrid 3
        if 2
        addr 10.10.3.254
/c/l3/vrrp/vr 3/track
                ports e
/c/l3/vrrp/vr 10
        ena
        vrid 10
        if 1
        addr 10.10.10.254
/c/l3/vrrp/vr 10/track
                ports e
/c/l3/rip/updat 30/on
/c/l3/rip/if 1/ena/supply e/listen e/default none/version 2
/c/l3/rip/if 1/poison d/split e/trigg e/metric 1/mcast e
/c/l3/rip/if 1/auth none
/c/l3/rip/if 2/ena/supply e/listen e/default none/version 2
/c/l3/rip/if 2/poison d/split e/trigg e/metric 1/mcast e
/c/l3/rip/if 2/auth none
/c/l3/rip/if 20/ena/supply e/listen e/default none/version 2
/c/l3/rip/if 20/poison d/split e/trigg e/metric 1/mcast e
/c/l3/rip/if 20/auth none
/c/l3/rip/if 25/ena/supply e/listen e/default none/version 2
/c/l3/rip/if 25/poison d/split e/trigg e/metric 1/mcast e
/c/l3/rip/if 25/auth none
/
script end  /**** DO NOT EDIT THIS LINE!

>> Configuration#
```

Example: A-16 C6506

```
c6506#sh run
Building configuration...

Current configuration : 2534 bytes
!
version 12.2
service timestamps debug datetime localtime
service timestamps log datetime localtime
no service password-encryption
service counters max age 10
!
hostname c6506
!
boot system flash
logging snmp-authfail
enable secret 5 $1$RZs1$2ck62Q8glJ1dk/jRYQaGr1
!
no aaa new-model
```

```
clock timezone EDT -4
ip subnet-zero
!
!
!
mls ip multicast flow-stat-timer 9
no mls flow ip
no mls flow ipv6
mls cef error action freeze
!
!
!
!
!
spanning-tree mode pvst
no spanning-tree optimize bpdu transmission
no spanning-tree vlan 15,20
!
spanning-tree mst configuration
 name TENGIG
 revision 1
 instance 3 vlan 3
 instance 10 vlan 10
!
diagnostic cns publish cisco.cns.device.diag_results
diagnostic cns subscribe cisco.cns.device.diag_commands
!
redundancy
 mode sso
 main-cpu
  auto-sync running-config
  auto-sync standard
!
vlan internal allocation policy ascending
vlan access-log ratelimit 2000
!
!
interface TenGigabitEthernet2/1
 no ip address
 switchport
 switchport trunk encapsulation dot1q
 switchport trunk allowed vlan 1-3
 switchport mode trunk
 switchport nonegotiate
!
interface TenGigabitEthernet2/2
 no ip address
!
interface TenGigabitEthernet2/3
 no ip address
 shutdown
!
interface TenGigabitEthernet2/4
 no ip address
 shutdown
```

```
!
interface TenGigabitEthernet3/1
 no ip address
 switchport
 switchport access vlan 15
 switchport trunk encapsulation dot1q
 switchport trunk native vlan 10
 switchport mode access
 switchport nonegotiate
!
interface TenGigabitEthernet3/2
 no ip address
 switchport
 switchport access vlan 20
 switchport trunk encapsulation dot1q
 switchport trunk native vlan 10
 switchport mode access
 switchport nonegotiate
!
interface TenGigabitEthernet3/3
 no ip address
!
interface TenGigabitEthernet3/4
 no ip address
!
interface GigabitEthernet6/1
 no ip address
 switchport
 switchport access vlan 3
 switchport trunk encapsulation dot1q
 switchport trunk native vlan 3
 switchport mode trunk
 switchport nonegotiate
 spanning-tree cost 10
!
interface GigabitEthernet6/2
 no ip address
 media-type rj45
 switchport
 switchport access vlan 3
!
interface Vlan1
 no ip address
 shutdown
!
interface Vlan15
 ip address 10.10.15.201 255.255.255.0
!
interface Vlan20
 ip address 10.10.20.201 255.255.255.0
!
interface Vlan30
 ip address 10.10.30.201 255.255.255.0
!
router rip
```

```
 version 2
 traffic-share min across-interfaces
 network 10.0.0.0
 network 192.168.1.0
!
ip classless
ip route 0.0.0.0 0.0.0.0 10.10.3.200
ip route 9.0.0.0 255.0.0.0 10.10.30.200
no ip http server
!
!
!
!
!
!
control-plane
!
!
!
line con 0
line vty 0 4
 password its0ral
 login
!
end
```

B

iSCSI NetApp configuration

This Appendix reviews the tests that were performed on iSCSI using 10 Gb networking from end to end.

B.1 Hardware

We tested a Network Appliance™ 7600 filer equipped with a 10 Gb NIC, attached by Multi-mode Fiber to a Nortel 10 Gb Ethernet Switch Module in a BladeCenter H chassis. In the BC-H chassis an HS21 server with a NetXen 10 Gb NIC was used.

B.2 Software

The filer was running a current Network Appliance version of their OnTAP operating system. The HS21 server was running Windows 2003 SP2. Version 3.4.216 of the NetXen drivers and firmware was used as was the case for most of our tests. A special version of the Microsoft iSCSI software initiator designed to support boot from iSCSI was tested in concert with IBM iSCSI Boot Commander. The latest version of iSCSI Boot Commander (2.04Boot) was used; this version of the initiator software can be accessed by getting the necessary invitation code from the IBM Web site.

B.3 Test results

In this section, we share our test results.

B.3.1 Data LUN Access

The server was able to successfully access a data LUN built on the filer. There were no observed differences in the configuration or behavior of the disk as accessed by the server or the filer. Benchmark tests were not performed as part of this exercise but other parties are

planning such tests in the near future. This test was essentially identical to the tests performed using 1 Gb connectivity to a Network Appliance FAS270 filer by using the MS software initiator described in the iSCSI IBM Redbooks publication. This book is available at: http://www.redbooks.ibm.com/abstracts/redp4153.html?Open.

B.3.2 Boot from iSCSI

We attempted to use the IBM iSCSI Boot Commander to test boot from iSCSI using the 10 Gb network infrastructure. We discovered that while the most recent version of the Boot Commander can detect the presence of the NetXen 10 Gb NIC, it does not support selecting that card to support iSCSI boot. It is not clear when or if a version of Boot Commander which does support the NetXen card will be available.

Future versions of the firmware for the NetXen card will support iSCSI boot by presenting the card to the blade's OS as an iSCSI HBA and a NIC at the same time. The timeframe for this function is not known at this time.

Abbreviations and acronyms

AMM	Advanced Management Module
BC-H	BladeCenter H chassis
BC-HT	BladeCenter HT Chassis
BCM	Broadcom
CF	Compact Flash
CFF-h	Combo Form Factor - Horizontal
CFF-v	Combo Form Factor - Vertical
CIM	Common Information Model
cKVM	Concurrent Keyboard Video and Mouse
CLP	Command Line Protocol
eth	Ethernet
ETSI	European Telecommunications Standards Institute
EXT	External
Gb	Gigabit
GEC	Gigabit EtherChannel
HDD	Hard Disk Drive
HPI	Hardware Platform Interface
HSDC	High-Speed Daughter Card
I/O	Input/Output
IBM	International Business Machines Corporation
INT	Interface
INT	Internal
IOS	Internetworking Operating System
ISL	Inter Switch Link
ITSO	International Technical Support Organization
LACP	Link Aggregation Control Protocol
LAG	Link Aggregation Group
LOM	LAN On Motherboard
LR	Long Reach/Long Range
MLT	Multi-Link Trunk
MM	Multi-mode
NEBS	Network Equipment-Building System
NEPS	Network Equipment Manufacturers
NIC	Network Interface Card
NX	NetXen
OSPF	Open Shortest Path First
RH	RedHat Linux
RHEL	RedHat Enterprise Linux
SFP	Small Form-factor Pluggable
SLES	SUSE Linux Enterprise Server
SM	Single-mode
SMASH	Systems Management Architecture for Server Hardware
SP	Service Provider
SR	Short Reach/Short Range
TOE	TCP Off-Load Engine
VLAN	Virtual LAN
XFP	10 Gigabit Small Form Factor Pluggable

Related publications

The publications listed in this section are considered particularly suitable for a more detailed discussion of the topics covered in this paper.

IBM Redbooks

For information about ordering these publications, see “How to get IBM Redbooks publications” on page 186. Note that some of the documents referenced here may be available in softcopy only.

- *Nortel Networks L2/3 Ethernet Switch Module for IBM eServer BladeCenter*, REDP-3586
- *Nortel Networks L2/3 GbESM Layer 2 High Availability Enhancements – L2/3 Switch Software Version 1.1*, TIPS0597
- *IBM BladeCenter Products and Technology, REDP-4264*
- *TCP/IP ports on the RSA II, BMC, and BladeCenter Management Module, TIPS0511*

Other publications

These publications are also relevant as further information sources:

- *Release Notes for Nortel 10 Gb Ethernet Switch Module,* 42c4910.pdf
- *Installation Guide for Nortel 10Gb Ethernet Switch Module,* 31r1751.pdf
- *Nortel 10Gb Ethernet Switch Module Application Guide,* 42c4911.pdf
- *Nortel 10Gb Ethernet Switch Module (Command Reference),* 42c4912.pdf
- *Nortel 10Gb Ethernet Switch Module (BBI Quick Guide),* 42c4913.pdf
- *ISCLI Reference Guide,* 42c4914.pdf
- *XFP Module option overview*, dw1cumst.pdf
- NetXen Installation and Tuning Guide

Online resources

These Web sites are also relevant as further information sources:

- Blade Network Technologies

  http://www.bladenetwork.net/pages/products_ibm_bladecenter.html
- NetXen

  http://www.netxen.com/index1.html
- IBM BladeCenter iSCSI SAN Solution

  http://www.redbooks.ibm.com/abstracts/redp4153.html?Open

How to get IBM Redbooks publications

You can search for, view, or download IBM Redbooks, Redpapers, Technotes, draft publications, and Additional materials, as well as order hardcopy IBM Redbooks, at this Web site:

ibm.com/redbooks

Help from IBM

IBM Support and downloads

ibm.com/support

IBM Global Services

ibm.com/services

Networking with the Nortel 10 Gb Ethernet Switch Module for BladeCenter H

Redpaper™

Demonstrates the capabilities of the Nortel 10 Gb ESM

Demonstrates the capabilities of the Nortel L2/3 10 Gb Uplink ESM

Highlights the NetXen NIC

This Redpaper discusses the physical layout of the 10 Gb switch and the installation in HSS slots 7 or 9 in BladeCenter H chassis. In addition, we discuss the physical layout of the NetXen HBA card and its installation in the HS21/LS21 and connectivity to an IBM TotalStorage device.

This Redpaper provides detailed instructions on how to install and configure NetXen drivers on the Windows 2K3 and Linux (RHEL and SUSE) platforms.

This Redpaper discusses:

- NIC Teaming and Bonding
- The 10 Gb switch configuration
  - Cisco integration, including Link Aggregation, multiple VLANs (802.1q), L3 routing with OSPF, Spanning Tree modes
  - HA features: Trunk failover, VRRP
  - Filters/access lists, including use of QoS; can include QoS example for iSCSI or other
- 10 Gb iSCSI
  - with N-Series/NetApp storage with 10 Gb NIC
  - Best performance with TCP Chimney or LSA

INTERNATIONAL TECHNICAL SUPPORT ORGANIZATION

BUILDING TECHNICAL INFORMATION BASED ON PRACTICAL EXPERIENCE

IBM Redbooks are developed by the IBM International Technical Support Organization. Experts from IBM, Customers and Partners from around the world create timely technical information based on realistic scenarios. Specific recommendations are provided to help you implement IT solutions more effectively in your environment.

For more information:
ibm.com/redbooks

REDP-4314-00