Redbooks Paper

Octavian Lascu
Pablo Pereira
Fernando Pizzano
Zbigniew Borgosz
Andrei Socolic
Josh-Daniel Davis

# IBM @server pSeries High Performance Switch Installation

For many years, IBM has been involved as a key player in High Performance Computing (HPC), introducing and developing technologies for implementing computing environments capable of processing massive amounts of data at very high speeds.

The main IBM technology for HPC was based on the concept of massively parallel processing (MPP). The practical implementation of this concept is the Scalable Parallel (SP) system, based on specialized hardware and managed by Parallel System Support Program (PSSP).

To deliver the performance required by the applications, in addition to powerful servers with high-speed processors, a low latency, very high bandwidth communication network was required. This type of network was originally implemented as the HiPS, evolving later to SP Switch, and then to SP Switch2.

When POWER4™ processor technology was available, a new type of switch was required to reap the full benefits of the capabilities of the new servers.

A new management infrastructure also had to be developed to deliver the functionality required by the new hardware and software stack. The new management function for the complex IBM @server pSeries® environment has been developed and named Cluster Systems Management (CSM). CSM provides the basic management functions required by a cluster based on either IBM @server pSeries or xSeries®, or a combination of the two server types.

Because CSM did not support existing SP Switch technology, a new switch technology was needed, the IBM @server pSeries High Performance Switch (HPS).

Although the SP Switch provides the structural basis of the IBM @server pSeries High Performance Switch (HPS), this particular iteration originated as a non-uniform memory access (NUMA) interconnect; therefore, NUMA requirements dictate its functions. NUMA requires a shared memory segment between two systems so that they can coordinate and operate as one. Hypervisor provides this function.

Although IBM has not announced any pSeries NUMA offering, the technological foundations remain, thereby providing a low-latency, high-speed, shared memory network.

Because NUMA requires its network to be active prior to the operating system, communication subsystem (CSS) support lacked the features required by the HPS. Due to the requirements of its original purpose, Hypervisor provides the functions needed by the HPS. Currently, only the pSeries 690 and 655 machines have the required resources to drive the HPS, and these machines must be in LPAR mode (in order to make use of Hypervisor, needed to manage the HPS).

## Naming the technology

Due to the rich history of the HPS, there are many names for its different components. We define the most common in Table 1.

*Table 1 HPS terminology*

| Term | Definition |
|---|---|
| fnmd | This is the daemon that manages topology, master selection, and initialization of the switch network. This performs its functions through hrdw_svr and presently runs on the HMC. |
| HMC | This is the Hardware Management Console. Also known as the Hardware Service Console. The Linux-based management station for LPAR-capable pSeries systems. |
| HPS | This stands for High Performance Switch. HPS is the shortest form of the official name IBM @server pSeries High Performance Switch. |
| LPAR | Logical partitioning. |
| pSeries HPS | See HPS. |
| SMA | Shared memory adapter, an obsolete name based on its technical function, can still be found in some reference materials. |
| SNI | Switch Network Interface is the current name for server-side components. |
| SNM | Switch Network Manager, used in the GUI to reference functions of the FNM. |

## Evolution of switch technology

The IBM @server pSeries High Performance Switch (HPS) is considered to be the next progression of the SP Switch. Although the HPS cannot be used with PSSP or an SP complex, its internal designs are similar to the SP Switch architecture. See Figure 1 on page 3 for the logical progression of IBM switch technology.

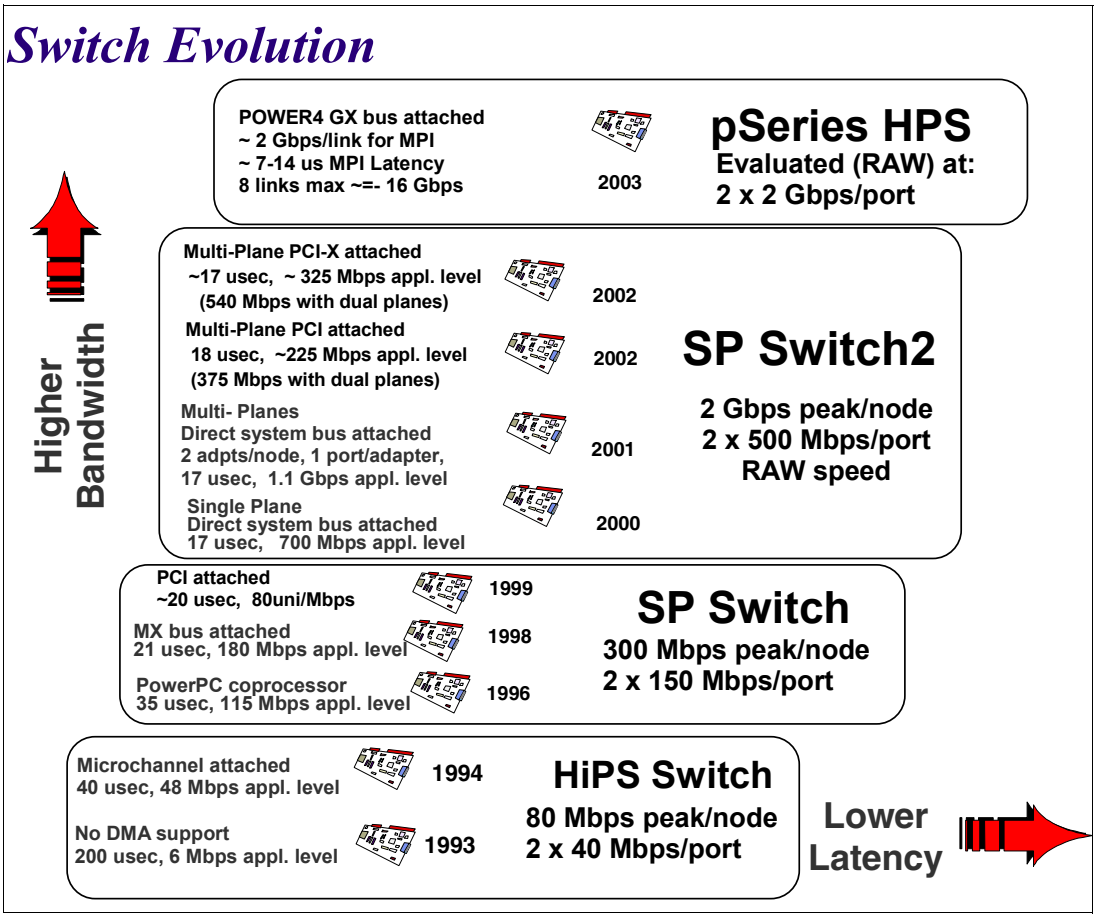


*Figure 1 IBM switch evolution*

# pSeries hardware requirements

As previously mentioned, the new HPS has special connection hardware requirements, thus the pSeries models supported for running the switch are limited.

The pSeries servers that currently support the HPS are:

- pSeries 690 (7040-681), running with CPU type:
  - POWER4 (1.1/1.3 GHz)
  - POWER4+™ (1.5/1.7 GHz)
- pSeries 655 (7039-651), running with CPU type:
  - POWER4 (1.1/1.3 GHz)
  - POWER4+ (1.5/1.7 GHz)

## Frames supported for HPS installation

The pSeries High Performance Switch hardware (M/T 7045-SW4) is supported for installation in the following specific frames:

- Maximum one switch per p690 (7040-681) frame

- Maximum one switch per pSeries 655 (7039-651) occupied frame (609.6 mm, 24 inch, wide, 42 EIA units, deep frame 7040-W42)
- One to eight switches in a 7040-W42 switch-only frame

If more than two switches are anticipated to be in the same switch-only frame, due to the bulk of the copper cables, add the 609.6 mm (24 in.) frame extender to your initial configuration to allow for proper clearance from the beginning.

For additional information, see the *IBM @server Cluster 1600 pSeries High Performance Switch Planning, Installation, and Service*, GA22-7951.

## Physical switch installation

The pSeries HPS is a four EIA-unit subsystem, so it occupies the same amount of rack space as an I/O drawer (M/T 7040-61D).

Within the initial offering, the primary components for an HPS are listed in Table 2.

*Table 2 Customer-orderable switch components*

| Feature code number | Description |
|---|---|
| FC 9049 | Intermediary Switch Board |
| FC 9047 | Server Switch Board |
| FC 6436 | Switch port connection card - optical (riser card) |
| FC 6435 | Switch port connection card - blank (slot filler) |
| FC 6433 | Switch port connection card - copper (riser card) |
| FC 3756 | Network diag tool (44P4060) (diag wrap SPCC, diag wrap copper cable, and diag wrap fiber cable) |

**Note:** The Intermediary Switch Boards (ISBs) come in sets of two or sets of four due to topology requirements.

The pSeries HPS has 16 slots used for installing the switch port connection cards (raiser cards):

- Eight dedicated slots for switch-to-node connections (each slot holds a 2-port riser card).
- Eight dedicated slots for switch-to-switch connections (each slot holds a 2-port riser card).

The slots for the switch-to-node and switch-to-switch riser cards are physically interleaved. An internal diagram of the HPS assembly (7045-SW4) is presented in Figure 2 on page 5.

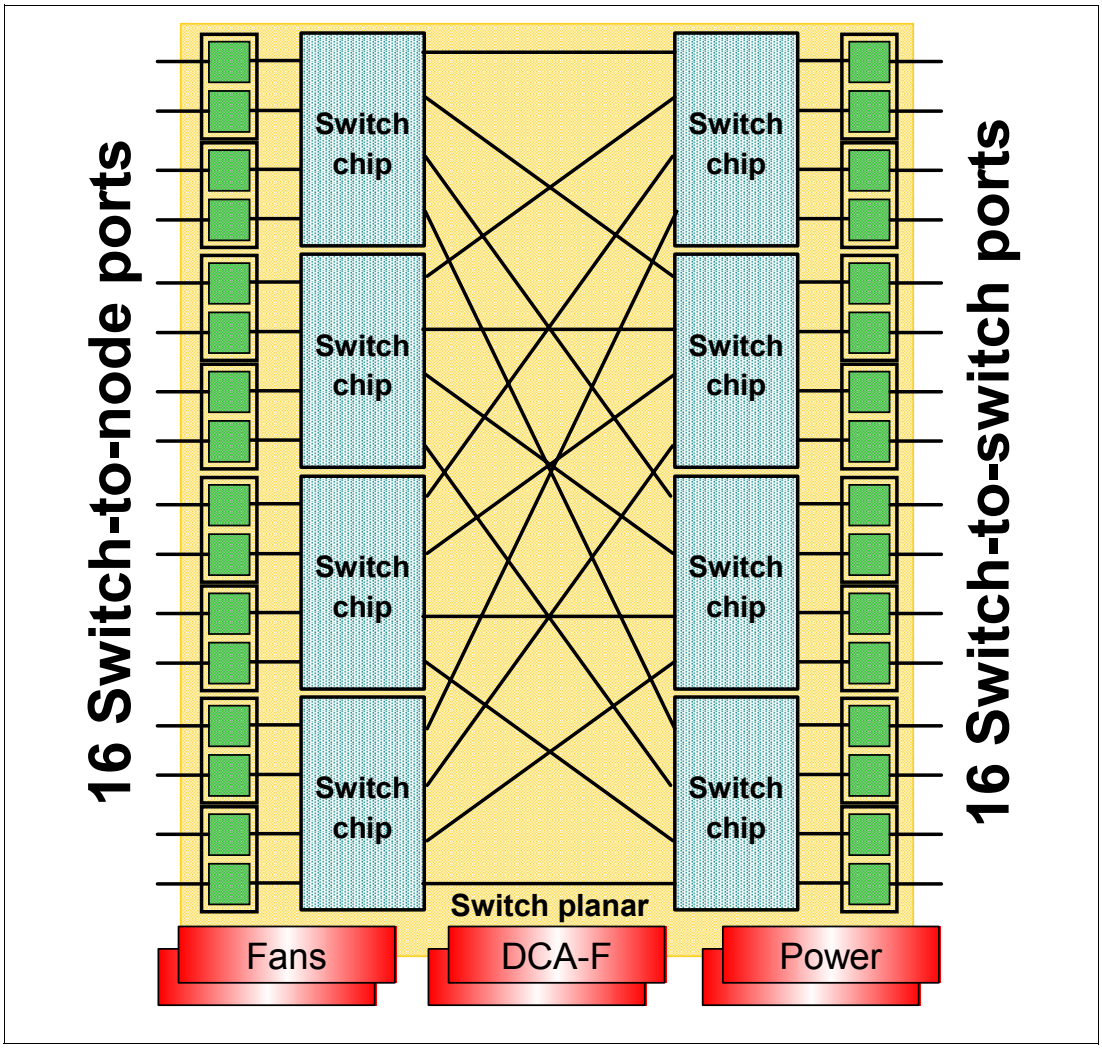


*Figure 2 HPS board: 1 plane, 32-port topology*

**Tip:** It is not mandatory to install the switch riser cards in sequential order. This might be useful in certain situations to provide a simplified switch cable management.

The switch riser cards provide the necessary electrical signal adaptation (transceivers) between the switch chip ports and cables. See Figure 4 on page 7.

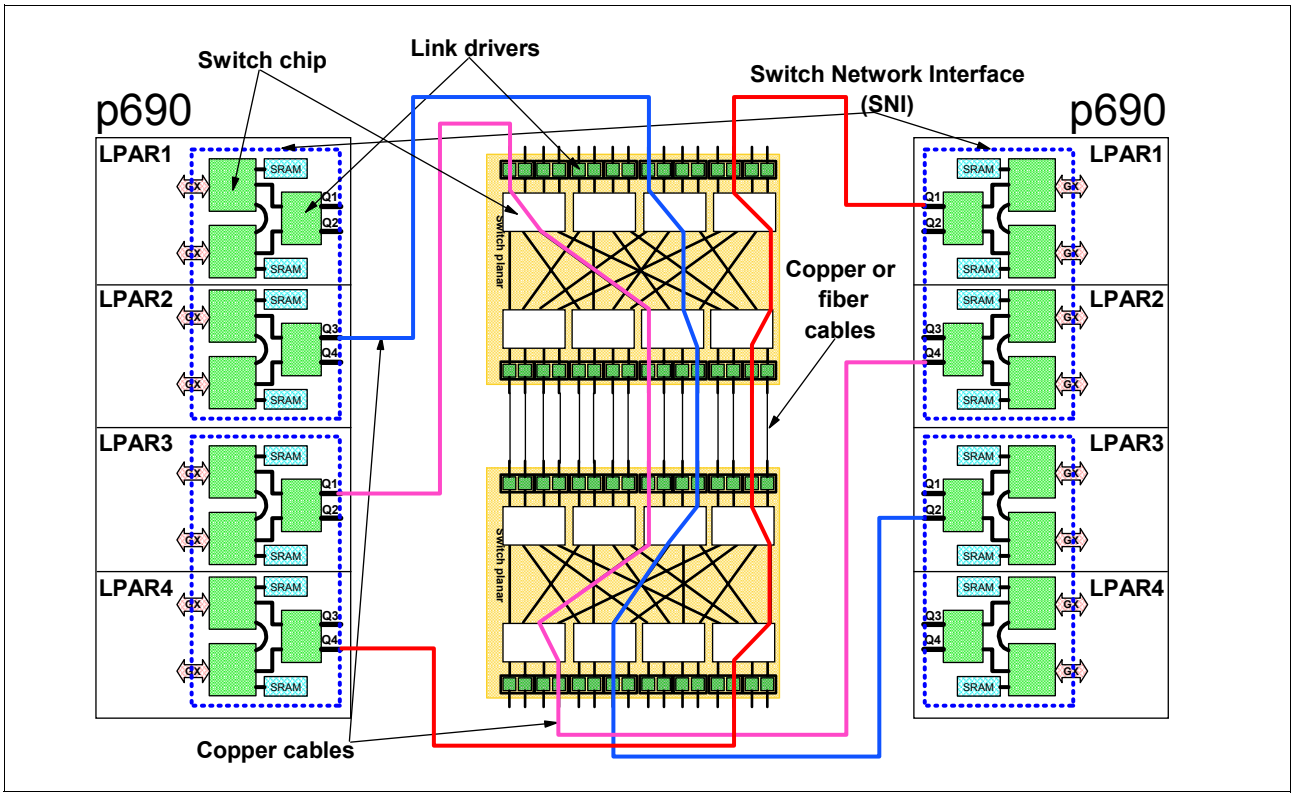


*Figure 3 Switch fabric components*

# Switch Network Interfaces

The system connects to the High Performance Switch using a Switch Network Interface (SNI) card attached to the GX bus slot (the SNI is the equivalent of a network interface card, or NIC). Depending on the pSeries system, the following Switch Network Interfaces are available:

- pSeries 690:
  - FC6434 IBM 4-Link Switch Network Interface for HPS
  - FC6432 IBM 2-Link Switch Network Interface for HPS
- pSeries 655:
  - FC6420 2-Link GX bus-mounted card

A diagram of a SNI card is presented in Figure 4 on page 7.

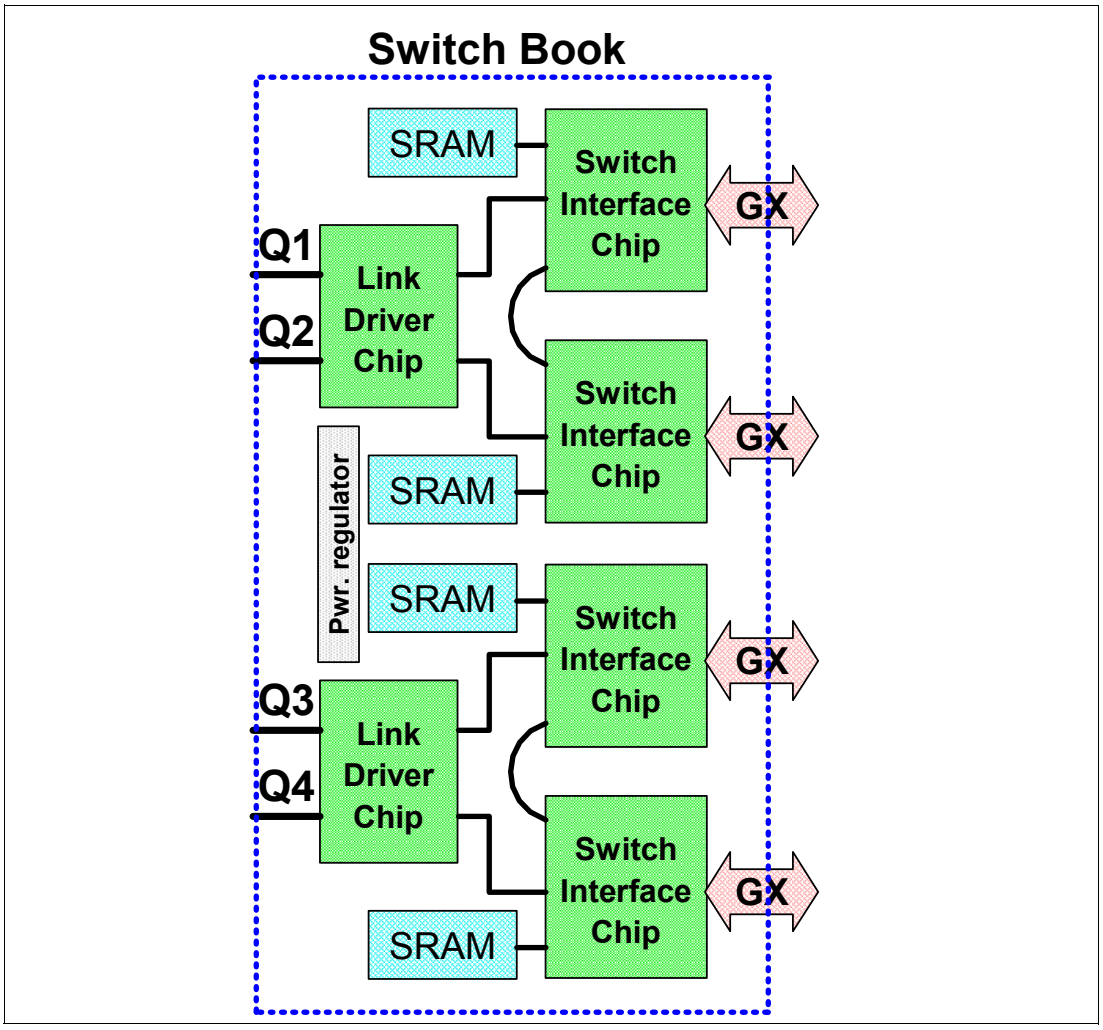


*Figure 4 Switch Network Interface card diagram*

For details about the GX slot locations, refer to *IBM @server Cluster 1600 pSeries High Performance Switch Planning, Installation, and Service*, GA22-7951.

The number of switch links per pSeries server depends on the type of server and its hardware configuration. For example, the number of supported links in a pSeries 690 system depends on the MCM configuration, and it might have 2, 4, 6, or 8 link pairs attached. See Figure 6 on page 9 for a correlation between the MCMs installed and the GX buses available in a pSeries 690 server.

## p655 SNI adapter placement

Due to bandwidth requirements, each SNI port requires its own GX bus. The p655 has only two GX buses available. Therefore, only one 2-Link adapter (FC 6420) can be used. Also, due to packaging and thermal considerations, the SNI card must be installed in the first GX slot; the second GX slot must remain empty.

*Table 3 Maximum links per pSeries 655 (7039-651) server for connecting to an HPS network*

| Server | Max. no. of links | Supported SNIs |
|---|---|---|
| p655 | 2 | 2-Link card (FC 6420) |

## p690 SNI adapter placement

The pSeries 690 (7040-681) servers provide one GX bus per CPU chip (2-way), for a total of four GX buses per MCM. Each GX slot provides two GX buses from each of its closest two MCMs, for a total of four GX buses per GX slot. Figure 5 presents the association between MCMs and GX slots.

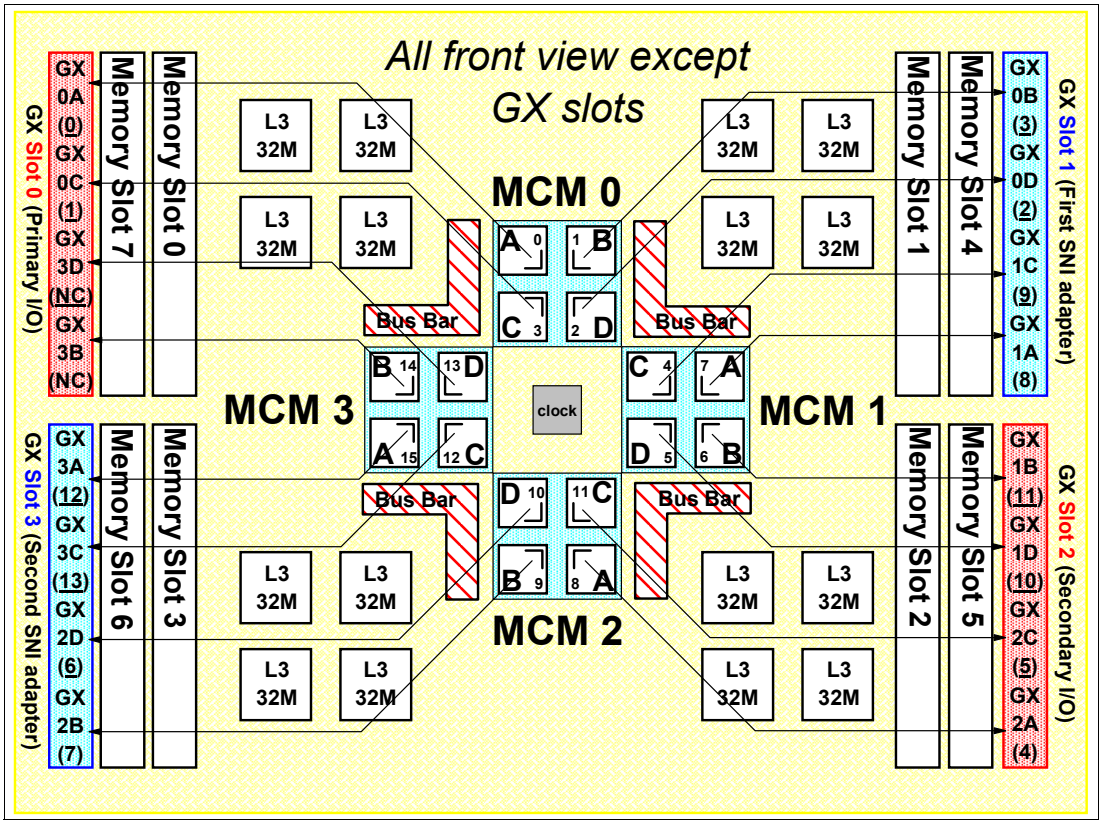


*Figure 5 p690 MCM to GX bus activation chart*

The number of switch links supported in a pSeries 690 server is shown in Table 4.

*Table 4 Maximum links per pSeries 690 (7040-681) server for connecting to an HPS network*

| Server | No. of MCMs | Max. no. of links | SNI configuration |
|---|---|---|---|
| p690 | 1 | 2 | 2-Link card (FC 6432) on GX slot 1 |
| p690 | 2 | 4 | 2-Link card (FC 6432) on GX slot 1<br>2-Link card (FC 6432) on GX slot 3 |
| p690 | 3 | 6 | 4-Link card (FC 6434) on GX slot 1<br>2-Link card (FC 6432) on GX slot 3 |
| p690 | 4 | 8 | 4-Link card (FC 6434) on GX slot 1<br>4-Link card (FC 6434) on GX slot 3 |

**Note:** The GX slots 0 and 2 are not supported for installing SNI books. Slot 0 is used for primary I/O book, and slot 2 for an additional I/O book.

The maximum configuration for a p690 (with four MCMs) is four LPARs connected to the HPS, each partition using two links. The links are always assigned in pairs. It is possible to

assign more than one link pair to the same LPAR for redundancy or for dual-plane configurations.

For example, in a pSeries 690 server configured with four MCMs in a single partition, all the SNIs (four link pairs) can be assigned to that partition. Figure 6 shows the relationship between the SNI books and the MCMs.

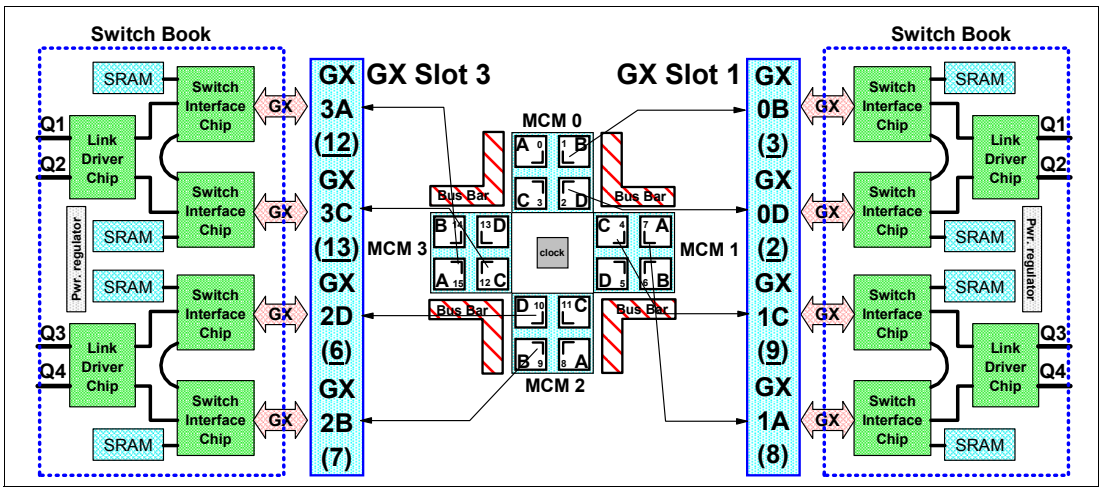


*Figure 6 MCM to GX to switch adapter relationship*

To allocate the SNI link pairs to the LPARs on a p690, see Figure 7.

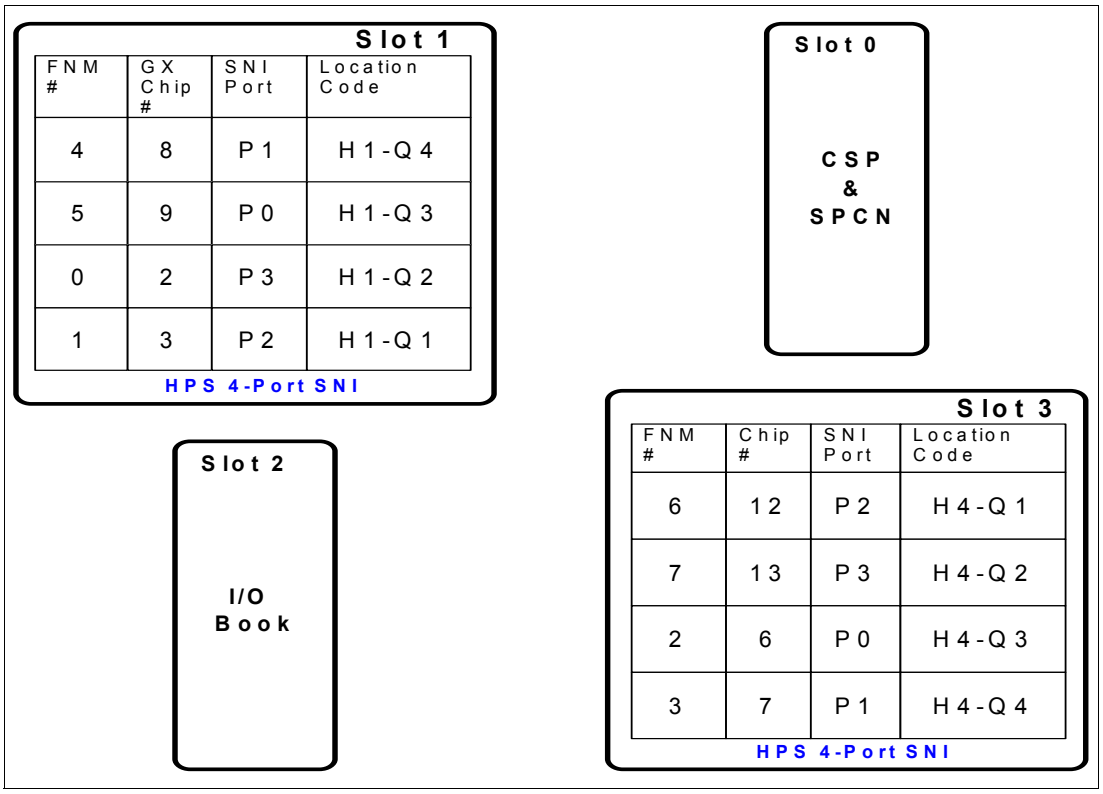


*Figure 7 SNI to GX bus map: Physical location codes*

Table 5 on page 10 describes the terminology used for physical location codes presented in Figure 6.

*Table 5 Physical location codes terminology*

| Term | Definition |
|---|---|
| SNI port | Described the port number within the book. |
| FNM # | Describes the network manager's logical designator for the chip connected to the port. |
| Chip # | The system's number used in CPU controls for that I/O chip based on the location on the GX bus. |
| Location code | Physical location of the slot and port within HMC I/O operations. |

### *Notes about adapter placement*

If a 4-port SNI adapter (book) is used where insufficient MCMs are available, all four links will be listed; however, inactive links assigned to an LPAR will prevent that LPAR from booting.

When both 4-port and a 2-port SNI books are used in the same Central Electronic Complex (CEC), the 4-port must be in slot 1 and the 2-port must be in slot 3.

## Installing the switch cables

There are two types of switch connections:

- Server-to-switch connections. This type of connection involves:
  - SNIs on the server side: 2-Link or 4-Link cards.
  - Copper cables switch port connection cards (FC 6433 - switch raiser cards) on the switch side.
  - Copper cables for connection between the SNI ports and the switch cards. The currently available cables for server to switch connection are:
    - Copper switch cable 1.2 m (FC 3161)
    - Copper switch cable 3 m (FC 3166)
    - Copper switch cable 10 m (FC 3167)

  The connections of an SNI link pair can be spread over the available switch cards in the same switch assembly, thus providing protection in case one switch raiser card fails.

  **Note:** The link pair provided by a 2-Link SNI can be converted to single-link use by installing a special warp plug (FC 6437). This feature is not supported for a 4-Link SNI card. Should this configuration become available, it will allow two LPARs to be connected to the same switch riser card. When using this configuration, 2-link nodes *cannot* span riser cards, and all cabling must be sequential.

- Switch-to-switch connections. This type of connection involves:
  - A pair of switch port connection cards, available in two options:
    - Copper switch port connection cards (FC 6433)
    - Fiber switch port connection cards (FC 6436)
  - Communication cables between switch port connection cards:
    - Copper cables: 1.2, 3, or 10 m
    - Multi-mode fiber cables: 20 m (FC 3256) or 40 m (FC 3257) cables

**Notes:**

- The same copper switch cards and cables are used for server-to-switch and switch-to-switch connections, but there are dedicated slots in the switch for each connection type. The fiber cards contains two links = two FC loops (IN, OUT pairs) for a total of four ports.
- A switch-to-switch connection must have both links of the switch riser cards connected.
- Single-link mode, as mentioned in *IBM @server Cluster 1600 pSeries High Performance Switch Planning, Installation, and Service*, GA22-7951, will not be a part of the initial offering.
- According to *IBM @server Cluster 1600 pSeries High Performance Switch Planning, Installation, and Service*, GA22-7951, a 4-Link SNI must be fully cabled, even if only one pair is used; however, we found that when cabling only one of the two pairs, it functions properly.
- According to *IBM @server Cluster 1600 pSeries High Performance Switch Planning, Installation, and Service*, GA22-7951, links within the same pair must attach to the same riser card.
- In a 2-switch, single-plane configuration, link pairs can be split between the two switches for redundancy.

## Installing the HMC administrative network

This sections describes how to install the Hardware Management Console (HMC) administrative network.

### Adapters for the local area network

You should plan for at least one network interface card (NIC) in each LPAR for the local area network (LAN) connection. For a IBM @server Cluster 1600, a LAN must connect the following resources:

- HMC
- LPARs
- Cluster Systems Management (CSM) server

The LAN adapters are used by the software functions included in the HMC code. They can be also used by the CSM server for management functions of LPARs and by the Network Install Manager (NIM) when performing a network operation against the LPARs.

Table 6 contains a list of HMC-supported Ethernet NICs.

*Table 6 Supported Ethernet adapters for HMC*

| Feature code number | Description |
|---|---|
| FC 4962 | 10/100 Mbps Ethernet PCI Adapter II |
| FC 2969 | Gigabit Ethernet SX PCI Adapter |
| FC 2975 | 10/100/1000 Base-T Ethernet PCI Adapter |

Currently, the supported speed for the trusted LAN is limited to 100 Mbps. The Ethernet adapters in a 7040-61D drawer for trusted and administrative LAN attachment must be installed in slots 8 and 9.

### Trusted network option for multiple HMCs

In a complex configuration, with multiple frames and SNM daemons running on multiple HMCs, a separate network for Switch Network Manager (SNM) communication should be provided. This is necessary because communication between SNM daemons running on separate HMCs should be not affected by any other network traffic (due to performance and security reasons). Therefore, we recommend using at least two network interfaces in each HMC, allocating one network interface for CSM-to-HMC communication, and one network interface for HMC-to-LPAR traffic.

**Tip:** This is not a technical requirement. In a small environment with a couple of HMCs and a few LPARs, both CSM-to-HMC and HMC-to-LPAR traffic can use the same physical and logical network.

## Installing hardware management cables

The HPS requires a mixture of RS-232 and RS-422 serial asynchronous ports in the HMC. For each CEC, three serial connections are required, one RS-232 connection to the Common Service Processor (CSP) and two RS-422 connections to the Bulk Power Assemblies (BPAs) in the frame. For a switch-only frame, only RS-422 connections are needed.

For a list of supported serial hardware, see Chapter 4, “System management Components,” in *IBM @server Cluster 1600 pSeries High Performance Switch Planning, Installation, and Service*, GA22-7951.

### CEC-to-HMC serial connection parameters

The hardware management connection between a Central Electronics Complex (CEC) and the HMC is made through an RS232 serial line running at 19200 bps, 8 data bits, no parity, 1 stop bit, and no hardware flow control lines. Framing is provided by Serial Line Internet Protocol (SLIP) within the Common Service Processor (CSP). This connection is now also used for setting up route tables within the Shared Memory Adapter (SMA), that is, the Switch Network Interface (SNI) controls.

### BPC-to-HMC serial communication parameters

The hardware management connection between the Bulk Power Assembly (BPA) and the HMC was previously provided through the CEC serial connection. Due to performance reasons, this has changed for frames that contain a pSeries HPS. The new connection is provided directly to the BPA through two RS422 serial lines running at 57600 bps, 8 data bits, no parity, 1 stop bit, balanced pairs for transmit and receive, and no hardware flow control lines. Framing is provided by Start of String Indicator (SOSI), an IBM proprietary protocol similar to SLIP. These new connections are used for frame power subsystem management such as firmware updates and switch hardware control.

For an example of serial cabling from HMC to frames, see Figure 8 on page 13.

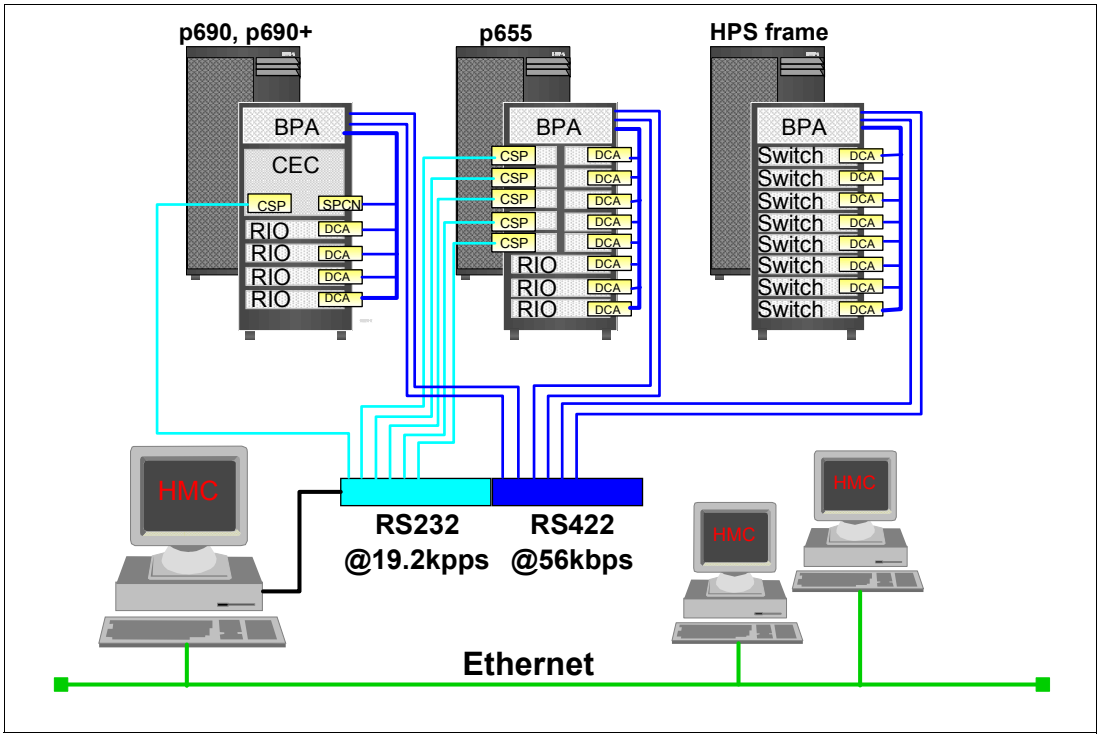


*Figure 8 Service network (multiple frames)*

**Note:** Although the 8-port asynchronous adapter natively supports both RS-232 and RS-422 modes, the 128-port asynchronous adapter only supports RS-232 through 16-port RANs. By the time the HPS is generally available (GA), an RS-232 to RS-422 convertor should be available as part of the RS-422 cable set as ordered.

### HMC requirements

The minimum configuration for the HMC should be a 7315-C01 or equivalent. The initial HPS offering is limited to 32 switch-attached nodes due to HMC performance considerations.

This limitation will likely be lifted by the next major release and might require new HMC hardware and software levels.

The version of the HMC code required for installing pSeries HPS is R3V2.4 or later. Check for the latest version of HMC code on the AIX® Support Web site:

`http://techsupport.services.ibm.com/server/hmc`

## Installing and configuring the pSeries High Performance Switch

This section describes the installation of pSeries HPS in our test environment. We also discuss different scenarios by providing general guidelines.

**Note:** The procedures related to HMC are provided for the hscroot user, unless a different authorization level is required for a specific operation.

## Sample test environment

For the examples presented in this section, we used the following configuration:

- Three IBM @server pSeries 690 systems (7040-681)
- One Hardware Management Console (7315-C01)
- Two IBM @server pSeries HPS (7045-SW4)

  The two switch assemblies are mounted in two separate pSeries 690 racks: 7040-61R.
- One IBM @server pSeries 630 (7028-6E4) as the CSM management server

Each of the p690 systems has 32 CPU and two 4-Link SNI cards. Four partitions have been created on each CEC, one LPAR containing eight CPUs and an SNI link pair. See Figure 9 for a picture of our LPAR environment as seen from the HMC.

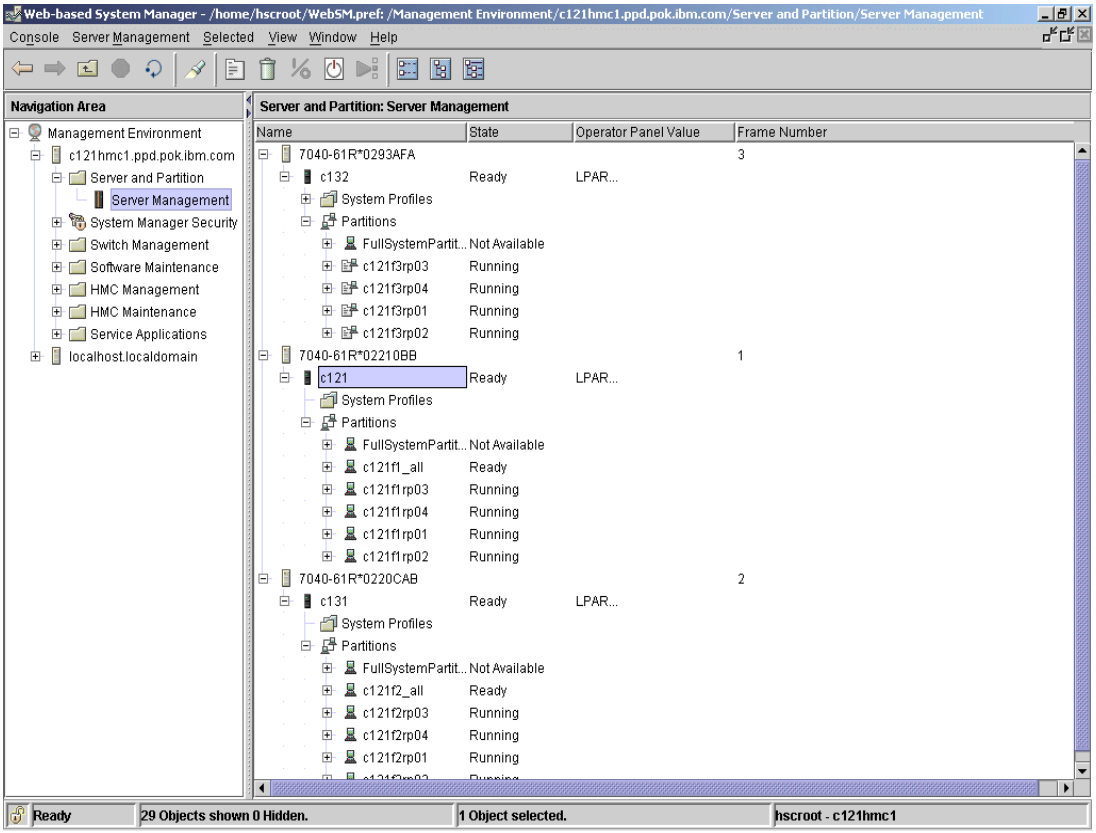


*Figure 9 Three CECs connected to pSeries HPS*

## Backing up the HMC configuration data

This section describes the methods used to save the critical data from your existing HMC. If you perform a new pSeries HPS installation with a fresh LPAR configuration, you can skip this step.

### Saving upgrade data

If you plan to upgrade existing HMC code, you have to save the data before starting the upgrade operation. The saved information includes:

- System preferences
- Profile information

- Service Agent files
- Inventory Scout Service files

The information can be stored either on a DVD-RAM media or on a separate partition on the HMC hard disk. This assumes that the hard drive will not be re-partitioned during the upgrade operation.

To save the upgrade data:

1. From the Navigation Area of the System Management interface, select **Software Maintenance**.
2. Select **HMC**.
3. Select **Save Upgrade Data**.
4. Select either **DVD** or **Hard drive**.

If you see an "HSCP0025" error message during the "Save Upgrade Data" operation, it might be due to insufficient space on the /dev/hda2 or /var file system. Check the /var/hsc/log/SaveUpgradeFiles.log file.

## Saving profile data

Use this function to save the LPAR profile information of your server.

To save the profile data:

1. In the Server Management window of the System Management interface, right-click the system icon of the targeted CEC for this operation.
2. Select **Profile Data**, and then **Backup**.
3. Enter the name of the backup file.

The file is saved in /var/hsc/profiles/*<MTMS>/<backup file name>*, where MTMS represents a string containing the machine type and the serial number of the CEC.

## Backing up critical console data

Using this task to save critical information about your HMC environment, such as:

- User preferences files
- HMC platform configuration files
- User information
- HMC log files

**Note:** This task requires a DVD-RAM media.

To back up critical console data:

1. From the Navigation Area of the Web-based System Manger (WebSM), select **Software Maintenance**.
2. Select **HMC**.
3. Select **Backup Critical Console Data**.

**Note:** If you need to restore your critical console data, you will need the HMC base media of the same version and release as the currently installed one. Restoring a backup on a different HMC version or release might result in a damaged HMC installation.

## Migrating the HMC code

To migrate the HMC code, first boot the HMC from the CD-ROM media.

At the Hardware Management Console, Hard Disk Install/Upgrade menu, you can choose:

1. Install/Recovery: For a new installation of the HMC code
2. Upgrade: For an upgrade installation of the new HMC code

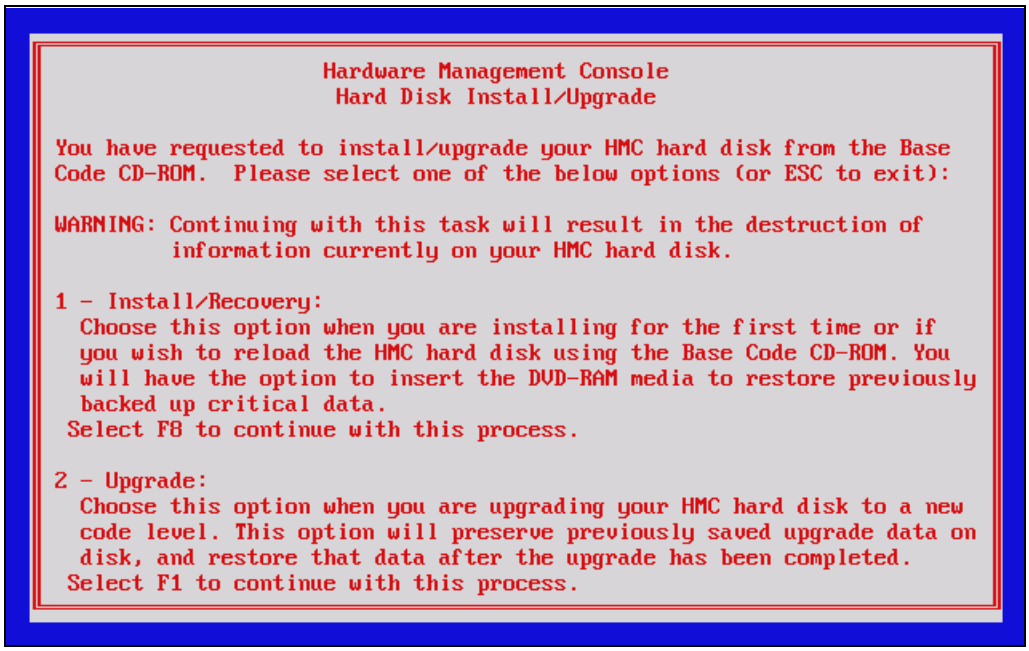


*Figure 10 Install/Upgrade HMC screen*

If you press F8 (Install/Recovery), a warning confirmation is issued (press F1 to continue). When this process ends, remove the CD from the DVD drive and press Enter.

If you press F1 (Upgrade), a confirmation is issued (press F1 to continue). When this process ends, remove the CD from the DVD drive.

If the HMC does not boot from the CD, press F1 when the HMC is booting to enter the BIOS Setup Utility, go to Startup Sequence, and check if the CD is on startup list.

## Configuring the HMC software

The following steps cover an installation of the HMC from scratch. If you upgraded your HMC, you should skip these steps.

### Changing the mouse, keyboard, and time zone configuration

If you are not using a USB mouse and keyboard, you need to change the mouse and keyboard configuration. Follow the on-screen instructions during the first reboot of the HMC after the installation process ends. You can also check the date, time, and time zone on the HMC. For further details, refer to *IBM @server Hardware Management Console for pSeries Installation and Operations Guide*, SA38-0590

## Configuring IP

To configure the IP:

1. From the System Manager interface on HMC, select **HMC Maintenance**.
2. Select **System Configuration**.
3. Select **Customize Network Settings**.

   Customize the network settings according to your environment. Some focus points:

| | |
|---|---|
| **IP address** | TCP/IP interface 0 address |
| | TCP/IP interface 0 network mask |
| | Default Gateway |
| **Name Services** | DNS Enable |
| | DNS Server Search Order |
| | Domain Suffix Search Order |
| **Host** | Host name |
| **Devices adapter** | Media speed for eth0 |

   **Note:** In order to activate the settings, reboot the HMC.

4. Test the network connectivity:

   a. Select **HMC Maintenance**.
   b. Select **System Configuration**.
   c. Select **Test Network Connectivity**.

   Provide a host name or an IP address of a valid host on the LAN (for example, your CSM server).

## Enabling the virtual terminal (optional)

You can enable the virtual terminal option on the HMC. This enables you to connect to the HMC using the Web-based System Manager Remote Client.

To enable the virtual terminal:

1. From the Web-SM Interface, select **HMC Maintenance**.
2. Select **System Configuration**.
3. Select **Enable** or **Disable Remote Virtual Terminal**. Then select **Enable Remote Virtual Terminal Connections** and click **OK**.

## Installing the HMC corrective service (if applicable)

This step installs the latest corrective services, available on the HMC Support Web site for your version of HMC code:

http://techsupport.services.ibm.com/server/hmc

To install the HMC corrective service:

1. From the System Management interface (that is, the WebSM), select **HMC Maintenance**.
2. Select **HMC**.
3. Install **Corrective Services**.

You can either choose a removable media or download the patches from a remote location.

### Configuring serial connectivity

You need to set up the serial adapters for RS232 and RS422 connections between the HMC and the managed systems.

To configure serial connectivity:

1. From the WebSM interface, select **HMC Maintenance**.
2. Select **System Configuration**.
3. Select **Configure Serial Adapter**.

   Select the number serial adapters you want to configure and their type. By default, the serial ports are configured like RS232 ports.
4. Reboot the system to activate the changes.

   After the reboot, continue with the configuration of the RS422 serial ports.
5. Select **HMC Maintenance**.
6. Select **System Configuration**.
7. Select **Configure Serial Adapter**.

   Select the board, port, and RS422 options.

   **Note:** If you have a 128-port asynchronous adapter with 16-port RANs, you do not have to configure the ports as RS422, because this function is provided by hardware convertors.

### Installation of Web-based System Manger remote client (optional)

If you want to manage the HMC from a remote workstation, you have to install the WebSM remote client on that workstation. Using the WebSM remote client, you can perform the same management operations as from the local HMC.

On your PC, open a browser window and go to `http://`*`HMC_name`*`/remote_client.html`, where HMC_name is the host name for you HMC.

On the Microsoft® Windows® NT or 2000 client, perform the following steps:

1. Save to disk.
2. Execute setup.exe.

The AIX WebSM interface (sysmgt.websm) also can serve the same purpose.

## Upgrading the firmware

For a pSeries HPS installation, you have to check that you have the correct version for two types of microcode:

- GFW (system firmware) is the microcode related to the CEC part of the system.

  The firmware image contains:

  - System power control network programming
  - Service processor programming
  - IPL programming
  - Run-time abstractization services

The minimum level of GFW required for connecting a system to the pSeries HPS is:

  - pSeries 690: 3H031024
  - pSeries 655: 3J031024

- Frame firmware is the microcode for the frame power subsystem. The minimum level required when installing a pSeries HPS is 259f.

Ensure that the latest firmware and microcode levels are installed on the system. The following firmware and microcode should be brought to the latest level:

- System microcode
- Frame firmware
- Integrated SCSI controller microcode
- Integrated Ethernet microcode

Because the SCSI and Ethernet controllers for the p690 system are actually placed on the I/O drawer, also check the firmware level associated with the I/O drawer 7040-61D.

For the latest firmware versions, check the IBM Technical Support Web site:

http://techsupport.services.ibm.com/server/mdownload2

System firmware should generally be loaded from floppy disk through the service processor menus; however, it can be loaded from AIX through normal system firmware methods.

Always check the release notes that come with the installation procedures, and update the system firmware to the level to correlate with the HMC software level. The first official HMC level supported for HPS installation is 3.2.5.

## Installing system firmware

The system firmware resides in the primary I/O book. We recommend that you perform this upgrade first. The minimum firmware version required for installing the pSeries HPS is:

- pSeries 690: 3H031024
- pSeries 655: 3J031024

There are several ways to update the system firmware, depending on the pSeries type. To download the latest firmware code and view the instructions for installing it, refer to:

http://techsupport.services.ibm.com/server/mdownload2

**Note:** Prior to an AIX installation, p690 firmware can be updated from service processor menus using a floppy disk drive. For p655, the microcode can be updated through NIM. Always check the release version and installation instructions accompanying the firmware.

For both p690 and p655, there is an alternate procedure for upgrading the system firmware by using an LPAR with AIX installed. This procedure assumes that:

- The LPAR to be used must have service authority.
- All other partitions except the one with service authority must be shut down.
- The partition with service authority must have the update image on a local file system.

To install the system firmware:

1. Log in as root to the LPAR with the service authority. All other LPARs must be shut down.

2. Create the directory /tmp/fwupdate if does not exist:

   ```
   mkdir /tmp/fwupdate
   ```

3. Download the proper firmware image from the IBM Support site, expand it, and put the image file in the /tmp/update directory.
4. Run these commands:

   ```
   cd /usr/lpp/diagnostics/bin
   ./update_flash -f /tmp/fwupdate/<image_file_name>
   ```

5. Follow the instructions on the screen. The procedure will shut down the system.

## Upgrading the frame microcode

**Note:** Before the installation of the frame firmware, make sure that the CEC is logically powered off (the LCD panel shows `0K`).

To upgrade the power subsystem microcode (applies to p690 only):

1. Disconnect the SPCN cables from the J00B (bottom) port on BPC A and BPC B.
2. Connect the RS-422 cables from the HMC to the J00B (bottom) port on BPC A and BPC B.

   **Note:** On the p690 server, the serial cables connecting the HMC to the BPC are normally connected to the J00A (top) port on the BPCs. However, the J00A ports must be enabled for use with the pSeries HPS. During this procedure, you will do the initial code load through the J00B ports. After you complete that process, you will move the RS-422 cables from the J00B ports to the J00A ports.

   Then, with the serial cables on the J00A ports, reinstall the microcode to enable the top ports for use with the pSeries HPS.
3. On the HMC System Management interface, select the **Software Maintenance** menu, and then perform the following steps:
   a. Select **Frame**.
   b. Select **Install Corrective Service**.
   c. Check the Frame code level (next to Frame MTMS in the lower panel).

      If the frame code level is 259f or higher, continue with step 4.

      If the frame code level is below 259f, do the following:

      i. Return to the Frame panel of the Software Maintenance menu.
      ii. Select **Receive Corrective Service**.
      iii. Place the disk with the ptcode into the HMC disk drive.
      iv. Click **OK** to upload from the disk.
      v. When the upload completes, return to the Frame panel of the Software Maintenance menu.
   d. Select **Install Corrective Service**.
   e. Highlight the Installed Version and the Frames to be updated fields. Highlight the appropriate corrective service version in the upper part of the Corrective Service panel, and select the frames to be updated in the lower part of the panel
   f. Select **Install** from the Install Corrective Service panel.

**Note:** The initial microcode install process can take up to one hour to complete. The second install (to enable the J00A ports) will only take a few minutes.

4. When the microcode installation completes, move the RS-422 cables from the J00B (bottom) ports to the J00A (top) ports on the BPCs.

   **Note:** If there are cables attached to the J00A ports, disconnect the cables from the J00A ports and label the cable ends as not used, or if possible, disconnect both ends of the cables completely and remove them from the frame.

5. Reattach the SPCN cables to the J00B ports.
6. Refresh the GUI to update the frame code display.
7. After you have attached the RS-422 cables to the J00A ports and the SPCN cables to the J00B ports, reinstall the microcode to enable the J00A ports for use with the pSeries HPS:
   a. Return to the **Software Maintenance** menu.
   b. Select **Frame**.
   c. Select **Install Corrective Service**.
   d. Highlight the Installed Version and the Frames to be updated fields. Highlight the appropriate corrective service version in the upper part of the Corrective Service panel, and select the frames to be updated in the lower part of the panel
   e. Select **Install** from the Install Corrective Service panel.
8. When the code is loaded and complete, power off the UEPO.

### *A note about Hypervisor*

Hypervisor is the glue that holds all of the HPS components together. It exists within the CEC as Licensed Internal Code and is part of the firmware image. When in LPAR Ready mode, Hypervisor owns all system resources and provides an abstraction layer through which device access and control is arbitrated. It is because of these functions that Hypervisor was chosen to handle the HPS.

The HPS was originally intended to be a NUMA fast-cache adapter. To do this, it needed to be available prior to system boot. This required much of the switch chip functionality to become part of system firmware, and only Hypervisor had the necessary hooks to provide abstraction of system resources. Although the NUMA on POWER4 plans have been discontinued, the HPS design has been continued. There is insufficient room in switch microcode to off load these functions from Hypervisor, and as such, the HPS requires the CEC to be in LPAR Ready mode.

We also recommend that you fully configure the Switch Network Manager prior to initial CEC power-on, after the addition of the switch. Failure to do so might prohibit the use of the switch due to improper frame identification within the hardware server, thereby crippling its ability to identify components within the switch network. SNM configuration can be done within the HMC GUI, or through /opt/hsc/data/HmcNetConfig.

## Configuring the frames

This section describes how to configure the frames.

### Switch Network Manager

Access the Switch Network Manager through the Switch Management GUI from within HMC V3.2.4 interface, as shown in Figure 11.

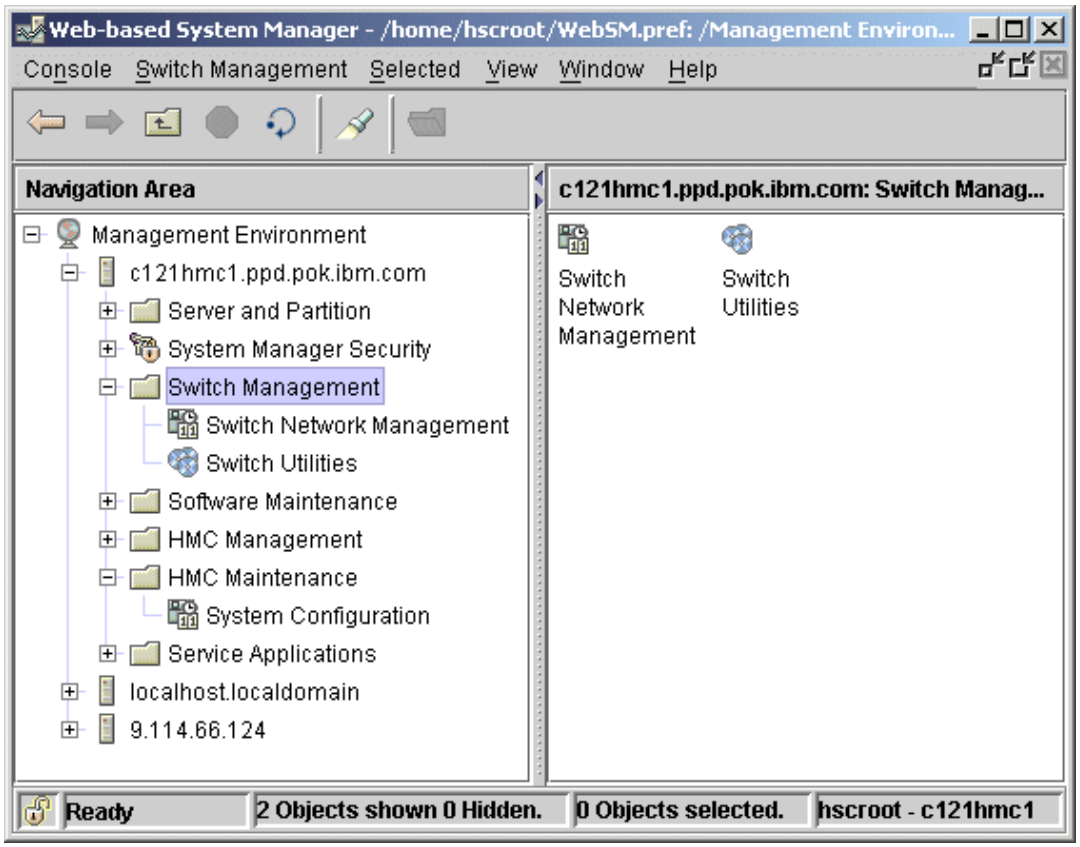


*Figure 11 Switch Management*

### Assigning frame numbers

After performing the frame power code upgrade and turning off the UEPO switch, you have to set the frame number to your frame.

On the HMC, perform the following steps:

1. Select **Switch Management**.
2. Select **Switch Utilities**.
3. Select **Frame Number Configuration**.
4. Set the frame number for the frame (see Figure 12 on page 23).

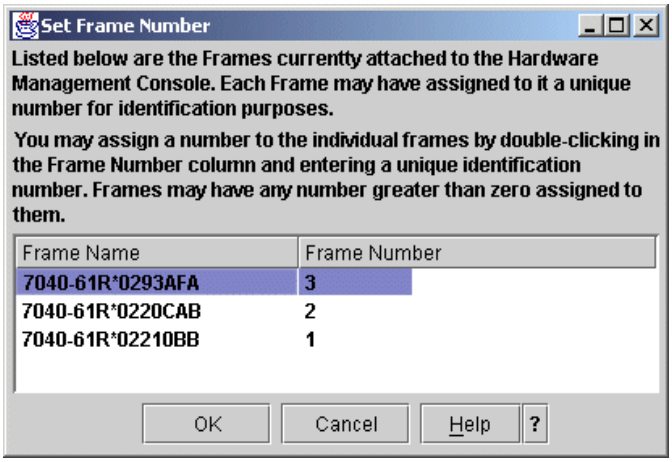


*Figure 12 Setting the frame number*

**Note:** HMC can manage both numbered and unnumbered frames. You can have a mixed environment containing switched and non-switched frames.

## Configuring LPARs

In order to use the HPS from AIX, you must assign links to your LPARs. Keep in mind that link pairs cannot be split between LPARs. Also, not all LPARs must be switch enabled if it does not meet your needs.

Links are always assigned in pairs and are not shared. This might change at some point, but currently, this is an architectural requirement. All internal components, including switch service processors and link pairs, are redundant. This means you can pull a switch cable loose, and rather than having an outage, a notice will show up in an AIX error report. We tested this, and only one ping packet was noticeably delayed.

If one link fails, the next available path is chosen (another link on the same SNI).

### *A note about affinity LPARs*

GX buses are I/O resources. I/O resources are not forcibly controlled in affinity partitions. Diagram Figure 5 on page 8 shows which MCMs are required for which GX buses, and Figure 7 on page 9 shows which GX buses are required for each link.

When assigning I/O to affinity LPARs, choose the link pairs attached to the CPUs that are assigned to your affinity LPAR. For details about how affinity LPARs' CPUs are chosen, see *IBM @server Hardware Management Console for pSeries Installation and Operations Guide*, SA38-0590.

**Restriction:** Due to reliance upon Hypervisor functions, switch adapters cannot be dynamically allocated (DLPAR).

## Creating a switch-enabled LPAR

To create a new LPAR attached to the switch:

1. Boot the CEC in LPAR mode:
   a. In the Server and Partition window, select the frame you want to start.
   b. Right-click the frame, and select **Power On**.

c. Select **Partition Standby mode**.

Wait for the LCD panel on HMC to display `LPAR`.

2. Select the frame on which you want to create the LPARs.
3. Right-click the frame, and select **Create** → **Logical Partition**.

   A new wizard starts to begin the configuration of the LPAR (see Figure 13). Enter a name for the partition, and click **Next**.

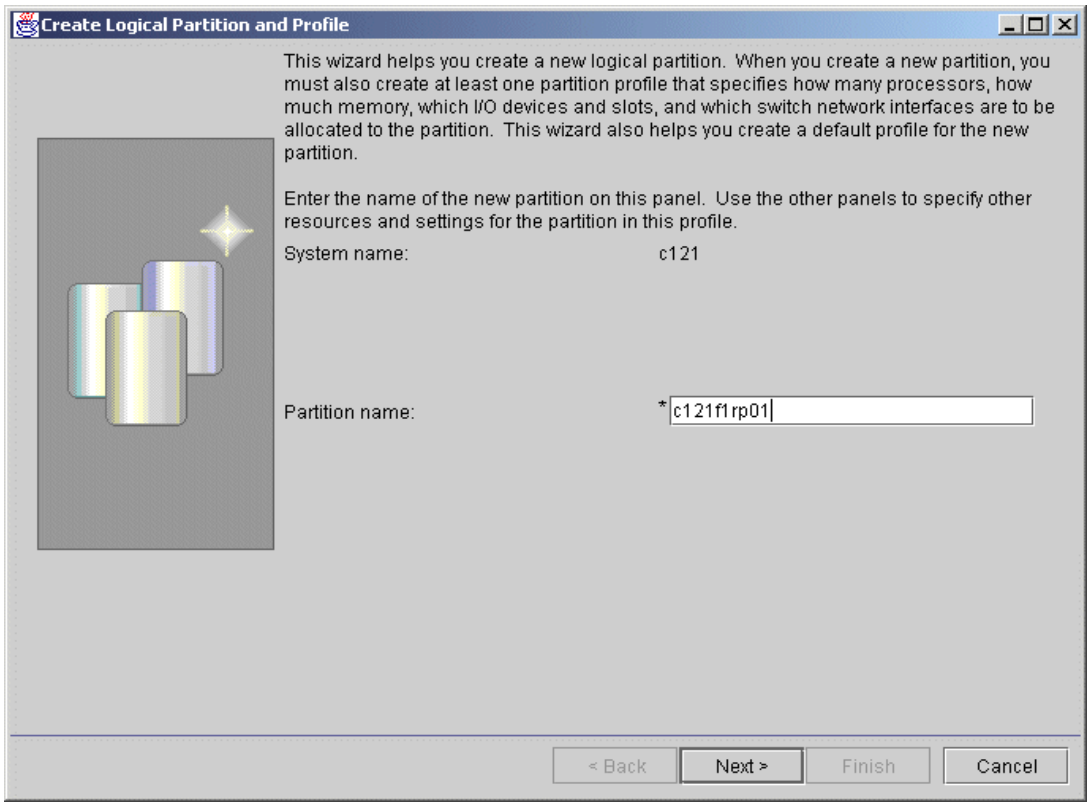


*Figure 13 Create a new LPAR*

4. Select the CPU, memory, and I/O resources for the new partition. After providing the CPU and memory resources, select the SNIs for your partition. See Figure 14 on page 25.

   For the p690 server used in our example, refer to the Figure 7 on page 9 for details related to the physical location of the SNI ports.

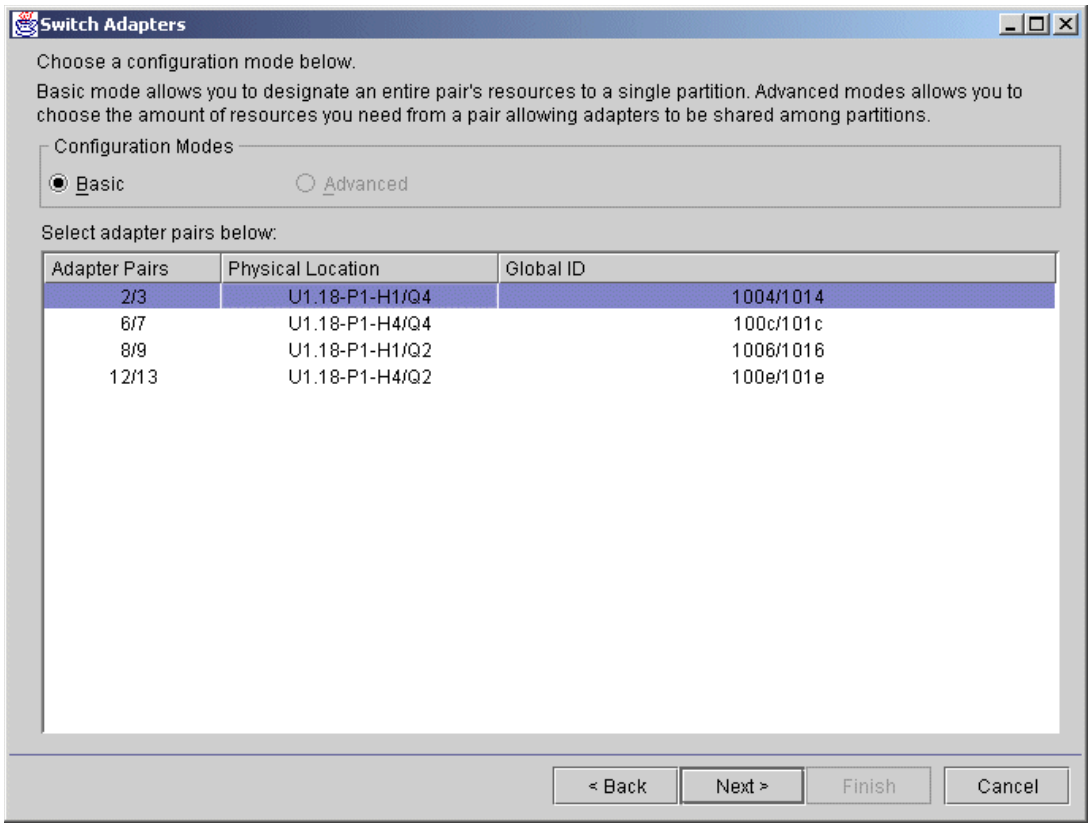


*Figure 14 Assign the SNIs to the LPAR*

**Notes:**

- Q2 and Q4 refer to the two link pairs of an 4-Link SNI adapter. The ports of an SNI adapter can be allocated to partitions only in pairs. A link pair cannot be split between partitions.
- The Global ID must not be empty; otherwise, you can allocate the link pair to the LPAR, but you will not be able to boot it. This can happen when a 4-port SNI card is installed in a system with only two MCMs.

## Upgrading an existing LPAR

If you are upgrading an existing LPAR environment for connection to the pSeries HPS, you should change the LPAR proprieties and add the desired SNIs to the LPAR.

To upgrade an existing LPAR:

1. From the Server and Partition window, expand the CEC icon containing the selected LPAR.
2. Right-click the LPAR name and select **Proprieties**.
3. Select the Switch Adapters tab and select the appropriate SNIs. See Figure 15 on page 26.

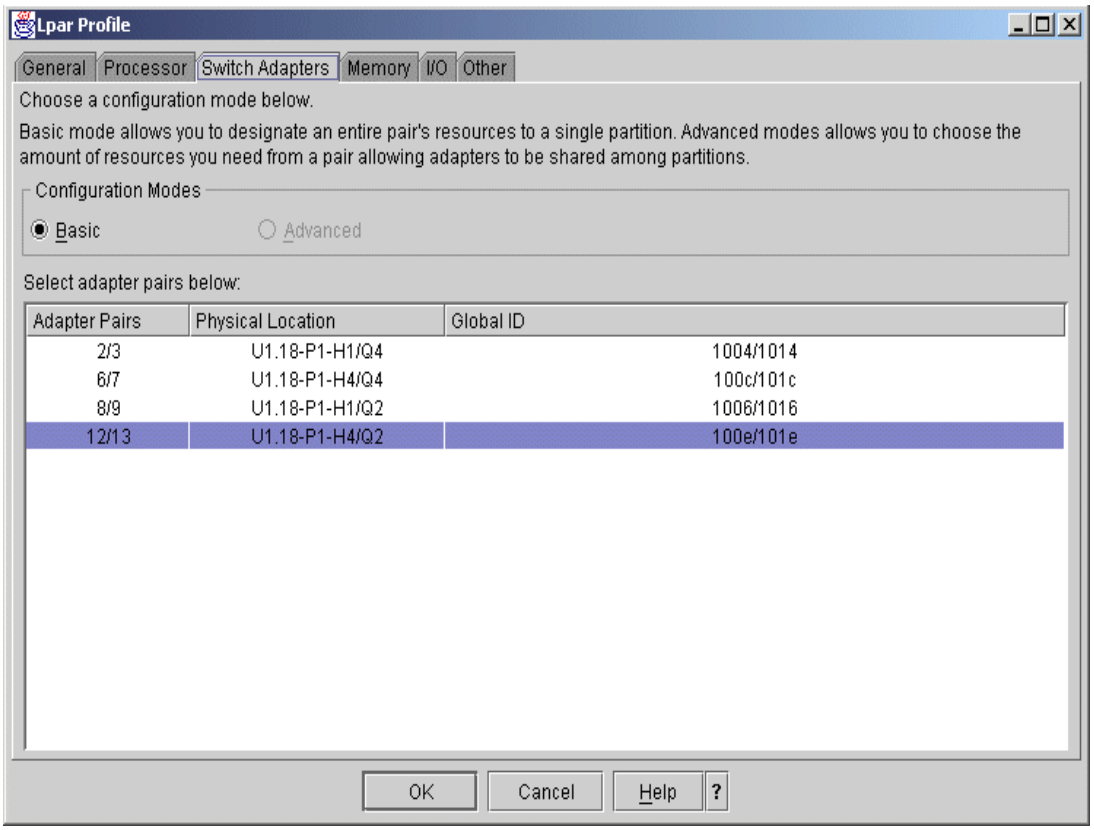


*Figure 15 Adding the SNIs to an existing LPAR*

**Note:** The SNIs cannot be dynamically allocated. For the changes to become effective, you must shut down or reset the partition and wait approximately 30 seconds prior to activation.

## Starting the switch

This section describes how to start the switch.

### Enabling and disabling SNM functions on the HMC

Perform the following steps when starting the switch for the first time:

1. Power off all the CECs using the HMC. Wait until "`No Connection`" and "`OK`" appears on the HMC console in Server Management window.
2. From the HMC WebSM interface:
   a. Select **Switch Management**.
   b. Select **Switch Network Management**.
   c. Select **Enable the SNM Software**. The window shown Figure 16 on page 27 opens. Click **Yes** in the confirmation window. This enables the FNM processes to start on the HMC. From this menu, you can also disable the FNM daemons by selecting **Disable the SNM Software**.

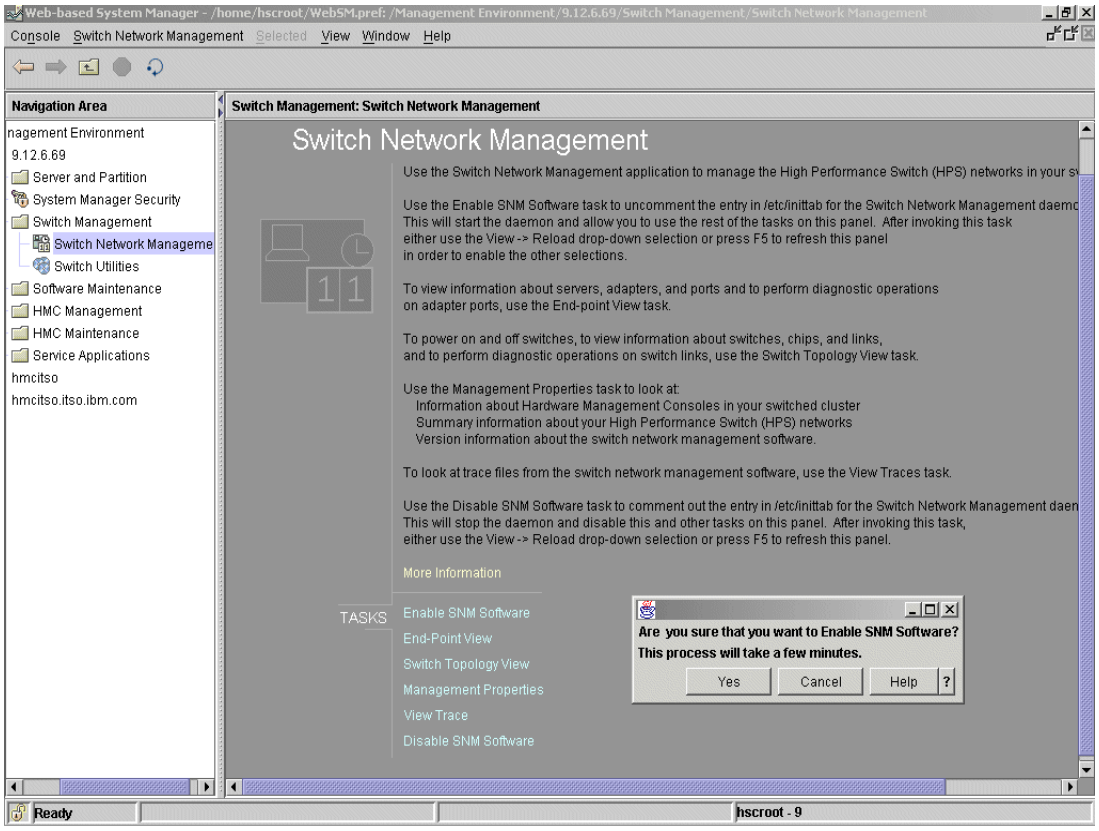


*Figure 16 Enable the Switch Network Management software*

3. Reboot the HMC.
4. Power on the CEC. Activate the Partition Standby Mode for booting the CEC in LPAR mode.
5. Check that the pSeries HPS is operational:
   a. Open a terminal window by pressing Ctrl-Alt-F1. You can switch back to the graphical mode by pressing Ctrl-Alt-F2.
   b. Switch from user “hscroot” to “root” and run the following command:

      ```
      /usr/local/hsctool/hps_check.pl
      ```

      Check the output of this command to ensure that it contains valid information (look at TIME, MPA, and TOD).

## Configuring CSM and NIM

This section describes the steps involved in configuring the Cluster Systems Management (CSM) environment. These steps are performed when the pSeries HPS is part of a Cluster 1600 environment. We recommend using CSM for installing and managing the LPARs connected to an HPS. This procedure assumes that you are adding new LPARs to an existing Cluster 1600 managed by CSM, so the CSM server and the Network Installation Management (NIM) environment are already initialized. For a new and complete installation of HMC, we recommend that you reboot the CSM server in this step.

## Adding CSM nodes

**Important:** Before CSM nodes are added to the CSM server, it is *mandatory* to ensure that name resolution is properly set up. If things mysteriously do not work, verify that the name resolution (IP address <-> IP label) gives identical results on the HMC, CSM master, and all HPS-attached LPARs. If DNS is used, verify that the /etc/resolv.conf file is identical among these machines. Forward and reverse host name lookups should be identical *for* all systems *from* all systems.

To add CSM nodes:

1. Enable access to the HMC. For the CSM server to be able to manage the LPARs, it must be able to access the HMC managing those LPARs. For our example, the HMC is `c121hmc1.ppd.pok.ibm.com` and the user used for access to the HMC resources is `hscroot`, as demonstrated in Example 1.

*Example 1 A systemid command example*

```
#systemid c121hmc1.ppd.pok.ibm.com hscroot
Password:
Verifying, please re-enter password:
systemid: Entry created.
```

2. Gather the hardware information about the LPARs managed by the HMC. Example 2 uses the **`lshwinfo`** command to collect information from the HMC (named c121hmc1) about the managed partitions.

*Example 2 Collecting the LPARs name on the CSM server*

```
#lshwinfo -p hmc -c c121hmc1.ppd.pok.ibm.com -o /tmp/c121hmc1.txt
#cat /tmp/c121hmc1.txt
# Hostname::PowerMethod::HWControlPoint::HWControlNodeId::LParID::HWType::HWModel::HWSerialNum
no_hostname::hmc::c121hmc1.ppd.pok.ibm.com::c121f2rp04::004::7040::681::0220CBB
no_hostname::hmc::c121hmc1.ppd.pok.ibm.com::c121f2rp03::003::7040::681::0220CBB
no_hostname::hmc::c121hmc1.ppd.pok.ibm.com::c121f1rp04::004::7040::681::0221OCB
no_hostname::hmc::c121hmc1.ppd.pok.ibm.com::c121f2rp02::002::7040::681::0220CBB
no_hostname::hmc::c121hmc1.ppd.pok.ibm.com::c121f3rp04::004::7040::681::0293B0A
no_hostname::hmc::c121hmc1.ppd.pok.ibm.com::c121f1rp03::003::7040::681::0221OCB
no_hostname::hmc::c121hmc1.ppd.pok.ibm.com::c121f2rp01::001::7040::681::0220CBB
no_hostname::hmc::c121hmc1.ppd.pok.ibm.com::c121f3rp03::003::7040::681::0293B0A
no_hostname::hmc::c121hmc1.ppd.pok.ibm.com::c121f1rp02::002::7040::681::0221OCB
no_hostname::hmc::c121hmc1.ppd.pok.ibm.com::c121f3rp02::002::7040::681::0293B0A
no_hostname::hmc::c121hmc1.ppd.pok.ibm.com::c121f1rp01::001::7040::681::0221OCB
no_hostname::hmc::c121hmc1.ppd.pok.ibm.com::c121f3rp01::001::7040::681::0293B0A
```

3. Edit the hostmap file and set the host name of the LPAR in the first field, as shown in Example 3.

*Example 3 The hostmap file after setting the host names*

```
# cat /tmp/c121hmc1.txt
# Hostname::PowerMethod::HWControlPoint::HWControlNodeId::LParID::HWType::HWModel::HWSerialNum
c121f2rp04::hmc::c121hmc1.ppd.pok.ibm.com::c121f2rp04::004::7040::681::0220CBB
c121f2rp03::hmc::c121hmc1.ppd.pok.ibm.com::c121f2rp03::003::7040::681::0220CBB
c121f1rp04::hmc::c121hmc1.ppd.pok.ibm.com::c121f1rp04::004::7040::681::0221OCB
c121f2rp02::hmc::c121hmc1.ppd.pok.ibm.com::c121f2rp02::002::7040::681::0220CBB
c121f3rp04::hmc::c121hmc1.ppd.pok.ibm.com::c121f3rp04::004::7040::681::0293B0A
c121f1rp03::hmc::c121hmc1.ppd.pok.ibm.com::c121f1rp03::003::7040::681::0221OCB
c121f2rp01::hmc::c121hmc1.ppd.pok.ibm.com::c121f2rp01::001::7040::681::0220CBB
c121f3rp03::hmc::c121hmc1.ppd.pok.ibm.com::c121f3rp03::003::7040::681::0293B0A
```

```
c121f1rp02::hmc::c121hmc1.ppd.pok.ibm.com::c121f1rp02::002::7040::681::0221OCB
c121f3rp02::hmc::c121hmc1.ppd.pok.ibm.com::c121f3rp02::002::7040::681::0293BOA
c121f1rp01::hmc::c121hmc1.ppd.pok.ibm.com::c121f1rp01::001::7040::681::0221OCB
c121f3rp01::hmc::c121hmc1.ppd.pok.ibm.com::c121f3rp01::001::7040::681::0293BOA
```

4. Define the nodes in CSM. The collected LPARs are added as endpoint nodes to the CSM server. Example 4 uses the LPAR definitions from Example 2 on page 28.

*Example 4 Defining LPARs as CSM-managed nodes*

```
#definenode -M /tmp/c121hmc1.txt
Defining CSM Nodes:
Defining Node "c121f2rp04.ppd.pok.ibm.com"("9.114.66.76")
Defining Node "c121f2rp03.ppd.pok.ibm.com"("9.114.66.75")
Defining Node "c121f1rp04.ppd.pok.ibm.com"("9.114.66.68")
Defining Node "c121f2rp02.ppd.pok.ibm.com"("9.114.66.74")
Defining Node "c121f3rp04.ppd.pok.ibm.com"("9.114.66.84")
Defining Node "c121f1rp03.ppd.pok.ibm.com"("9.114.66.67")
Defining Node "c121f2rp01.ppd.pok.ibm.com"("9.114.66.73")
Defining Node "c121f3rp03.ppd.pok.ibm.com"("9.114.66.83")
Defining Node "c121f1rp02.ppd.pok.ibm.com"("9.114.66.66")
Defining Node "c121f3rp02.ppd.pok.ibm.com"("9.114.66.82")
Defining Node "c121f1rp01.ppd.pok.ibm.com"("9.114.66.65")
Defining Node "c121f3rp01.ppd.pok.ibm.com"("9.114.66.81")

#lsnode
p690_LPAR1.itso.ibm.com
p690_LPAR2.itso.ibm.com
c121f2rp04.ppd.pok.ibm.com
c121f2rp03.ppd.pok.ibm.com
c121f1rp04.ppd.pok.ibm.com
c121f2rp02.ppd.pok.ibm.com
c121f3rp04.ppd.pok.ibm.com
c121f1rp03.ppd.pok.ibm.com
c121f2rp01.ppd.pok.ibm.com
c121f3rp03.ppd.pok.ibm.com
c121f1rp02.ppd.pok.ibm.com
c121f3rp02.ppd.pok.ibm.com
c121f1rp01.ppd.pok.ibm.com
c121f3rp01.ppd.pok.ibm.com
```

5. Create the node groups. In Example 5, we create three node groups associated with the CEC frames of the nodes.

*Example 5 Creating node groups*

```
# nodegrp -n c121f1rp01.ppd.pok.ibm.com,c121f1rp02.ppd.pok.ibm.com,c121f1rp03.ppd.pok.ibm.com, \
c121f1rp04.ppd.pok.ibm.com frame1_grp
# nodegrp -n c121f2rp01.ppd.pok.ibm.com,c121f2rp02.ppd.pok.ibm.com,c121f2rp03.ppd.pok.ibm.com, \
c121f2rp04.ppd.pok.ibm.com frame2_grp
# nodegrp -n c121f3rp01.ppd.pok.ibm.com,c121f3rp02.ppd.pok.ibm.com,c121f3rp03.ppd.pok.ibm.com, \
c121f3rp04.ppd.pok.ibm.com frame3_grp
# nodegrp | grep frame
frame1_grp
frame2_grp
frame3_grp
```

## Installing AIX on the LPARs

Only AIX 5.2.0.0-ML02 with APAR IY49612 or later is needed to use the IBM @server pSeries HPS. NIM now includes a feature known as the secondary NIM adapter that allows additional adapters to be configured (similar to PSSP functionality). Refer to “Installing and configuring the SNIs in AIX” on page 33.

There is no support for Linux on pSeries.

### *Device support*

The switch drivers are now a part of AIX and are included in the filesets shown in Example 6. Links within an SNI, link devices within AIX, and protocol devices within AIX are all managed by the device HPS device drivers.

*Example 6 SNI driver filesets for V1.1.0.1*

```
# installp -ld .
devices.common.IBM.sni.rte 1.1.0.1 Switch Network Interface Runtime
devices.common.IBM.sni.ml 1.1.0.1 Multi Link Interface Runtime
devices.common.IBM.sni.ntbl 1.1.0.1 etwork Table Runtime
devices.chrp.IBM.HPS.rte 1.1.0.1 IBM eServer pSeries High Performance Switch (HPS) Runtime
devices.chrp.IBM.HPS.hpsfe 1.1.0.1 IBM pSeries HPS Functional Exerciser
devices.msg.en_US.common.IBM.sni.rte 1.1.0.1 Switch Network Interface Rte Messages
devices.msg.en_US.common.IBM.sni.ml 1.1.0.1 Multi Link Interface Runtime
devices.msg.en_US.common.IBM.sni.ntbl 1.1.0.1 Network Table Runtime Messages
devices.msg.en_US.chrp.IBM.HPS.rte 1.1.0.1 pSeries HPS runtime Messages
devices.msg.en_US.chrp.IBM.HPS.hpsfe 1.1.0.1 pSeries HPS functional utility
```

Updates to AIX filesets are generally available at:

https://techsupport.services.ibm.com/server/aix.fdc

**Important:** The SNI device drivers require the 64-bit kernel. The kernel fileset required for the installation in a pSeries HPS environment must be at least bos.mp64.5.2.0.14. This fileset is included in the AIX 5L™ V5.2 ML2 package.

## Checking the rpower command

Example 7 illustrates the **rpower** command.

*Example 7 rpower command example*

```
# rpower -N frame1_grp,frame2_grp,frame3_grp query
c121f2rp04.ppd.pok.ibm.com off
c121f2rp03.ppd.pok.ibm.com off
c121f1rp04.ppd.pok.ibm.com off
c121f2rp02.ppd.pok.ibm.com off
c121f3rp04.ppd.pok.ibm.com off
c121f1rp03.ppd.pok.ibm.com off
c121f2rp01.ppd.pok.ibm.com off
c121f3rp03.ppd.pok.ibm.com off
c121f1rp02.ppd.pok.ibm.com off
c121f3rp02.ppd.pok.ibm.com off
c121f1rp01.ppd.pok.ibm.com off
c121f3rp01.ppd.pok.ibm.com off
```

## Getting adapter MAC addresses

LPAR adapters will be configured by NIM using information from CSM. Example 8 shows a method for getting the MAC address of the node c121f1rp01.ppd.pok.ibm.com® using the **getadapters** command and storing it in CSM. You can choose to add the MAC address of your node manually if it is available.

**Important:** In CSM Version 1.3.2.1, the **getadapters** command first tries to establish a remote shell connection to the node only if that node has the attribute Mode set to Managed or MinManaged. If the remote shell connection fails, the **getadapters** command tries to use the hardware control method, which might result in rebooting your LPAR at the open firmware prompt (assuming hardware control is available).

*Example 8 Gathering adapter information in CSM*

```
# getadapters -w -t ent -D -s 100 -d full -n c121f1rp01.ppd.pok.ibm.com
# Name::Adapter Type::MAC Address::Location Code::Adapter Speed::Adapter Duplex::Install Server::Adapter
Gateway::Ping Status
Acquiring adapter information using dsh.
Can not use dsh - No nodes in Managed or MinManaged mode.
Acquiring adapter information from Open Firmware for node c121f1rp01.ppd.pok.ibm.com.
c121f1rp01.ppd.pok.ibm.com::ent::0002556A5352::U1.9-P1-I3/E1::100::full::csm_server.ppd.pok.ibm.com::0.0.0.
0::ok
# lsnode -l c121f1rp01.ppd.pok.ibm.com|grep InstallAdapter
 ChangedAttributes = {InstallAdapterMacaddr,InstallAdapterType,InstallAdapterSpeed,InstallAdapterDuplex}
 InstallAdapterDuplex = full
 InstallAdapterGateway =
 InstallAdapterMacaddr = 0002556A5352
 InstallAdapterNetmask =
 InstallAdapterSpeed = 100
 InstallAdapterType =
```

In Example 8, the installation adapter is set to 100_full_duplex mode. You should choose the settings according to your network environment. You can check the adapter attributes in the CSM node definition using the **lsnode** command and make the necessary changes if needed.

## Configuring the NIM environment

To configure the NIM environment:

1. Create the NIM resources:
   a. Create the LPP and shared product object tree (SPOT) resources.

      The lppsource must contain the AIX 5L V5.2 package and the patches for ML02. In addition, apply the SNI filesets to the lppsource and create the SPOT, as described in Example 9 on page 32.

*Example 9 Creating the lppsource and SPOT*

```
# echo Create the lppsource_52ML2 resource
# nim -o define -t lpp_source -a source=/dev/cd0 -a server=master \
-a location=/csminstall/AIX/aix520/lppsource lppsource_52ML2
# echo Update the lppsource with the ML02 package from a local directory: /tmp/AIXML02
# nim -o update -a packages=all -a source=/tmp/AIXML02 lppsource_52ML2
# echo Update the lppsource with the SNI file sets from directory: /tmp/SNI
# nim -o update -a packages=all -a source=/tmp/SNI lppsource_52ML2
# echo Create the SPOT
# nim -o define -t spot -a source=lppsource_52ML2 -a server=master \
-a location=/csminstall/export/ spot52_ML2
```

Example 9 assumes that the AIX package is loaded from the CD-ROM media. The AIX patches and the SNI filesets are downloaded to local directories. For a list of SNI filesets, refer to Example 6 on page 30.

b. Create and customize a bosinst_data resource.

You can use the template file /usr/lpp/bosinst/bosinst.template and customize it.

**Important:** You must use ENABLE_64BIT_KERNEL = yes in the stanza file to activate the 64-bit kernel required by the SNI drivers. Otherwise, the SNIs might not be activated.

For creating the bosinst_data resource, see Example 10.

*Example 10 Create a bosinst_data resource*

```
# nim -o define -t bosinst_data -a server=master -a location=/csminstall/bosinst.data bosinst_data_52
```

c. Define a resolv_conf resource.

You can define a resolv_conf resource in NIM for setting the name resolution configuration for the nodes. Example 11 shows the creation of a resource named resolv_conf_52, using the /csminstall/resolv.conf file.

*Example 11 Define a resolv_conf resource*

```
# nim -o define -t resolv_conf -a server=master -a location=/csminstall/resolv.conf resolv_conf_52
```

d. Define a default route for the NIM environment.

You can define a default route for your environment on the NIM network resource. See Example 12, where the nim_network is the primary network, defined when the NIM master was initialized.

*Example 12 Define a default route in the nim_network resource*

```
#nim -o change -a routing1="default 9.144.66.1" nim_network
```

2. Create a NIM node environment from CSM. At this step, use the csm2nimgrps and csm2nimnodes scripts provided by CSM to create the NIM nodes and groups from the CSM definitions. In Example 13 on page 33, we use the nodes and groups previously created in Example 4 on page 29 and Example 5 on page 29.

*Example 13 Create the NIM nodes and node groups*

```
# echo Create NIM node objects
# csm2nimnodes -N frame1_grp,frame2_grp,frame3_grp type=standalone platform=chrp netboot_kernel=mp \
network_name=nim_network cable_type="N/A"
# echo Create NIM node groups
# csm2nimgrps -N frame1_grp,frame2_grp,frame3_grp
# echo Check that the NIM definitions were created:
# lsnim
```

**Note:** The csm2nimnodes script requires the CSM nodes to have the network attributes defined in CSM.

3. Set up the CSM client software to be installed on the nodes. See Example 14.

*Example 14 Setting the CSM client to be installed on the nodes*

```
#csmsetupnim -N frame1_grp,frame2_grp,frame3_grp
```

Use the **lsnim** command to check that a new resource called csmprereboot_script was created.

4. Set up the nodes to be installed from the network. See Example 15.

*Example 15 Set up nodes to be installed from the network*

```
# echo Allocating the NIM resources to nodegroups
# nim -o allocate -a spot=spot52_ML2 -a lpp_source=lppsource_52ML2 -a bosinst_data=bosinst_data_52 \
-a resolv_conf=resolv_conf_52 -a script=csmprereboot_script frame1_grp
# nim -o allocate -a spot=spot52_ML2 -a lpp_source=lppsource_52ML2 -a bosinst_data=bosinst_data_52 \
-a resolv_conf=resolv_conf_52 -a script=csmprereboot_script frame2_grp
# nim -o allocate -a spot=spot52_ML2 -a lpp_source=lppsource_52ML2 -a bosinst_data=bosinst_data_52 \
-a resolv_conf=resolv_conf_52 -a script=csmprereboot_script frame3_grp

# echo Setting nodes for BOS installation on netboot
# nim -o bos_inst -a source=rte -a boot_client=no -a accept_licenses=yes frame1_grp
# nim -o bos_inst -a source=rte -a boot_client=no -a accept_licenses=yes frame2_grp
# nim -o bos_inst -a source=rte -a boot_client=no -a accept_licenses=yes frame3_grp
--
```

5. Network boot the target nodes.

In Example 16, we install the nodes in groups. Issue a **netboot** command for booting the nodes from the network and the **rconsole** command for opening a console and monitoring the installation process.

*Example 16 Network boot a group of nodes*

```
# netboot -N frame1_grp
```

For opening a console to monitor the installation, use the **rconsole -n node_name** command.

You have to run this command from the graphical console.

## Installing and configuring the SNIs in AIX

There are no more “E-commands,” and there is no fault service daemon or rc.switch. As such, regular AIX **ifconfig** and **smitty tcpip** can be used to bring up and down the SNI (sn#) interfaces. The logical device sn# refers to the logical device name of one of the external links on an SNI. For example, if allocated to a single LPAR, the 4-Link SNI book will

present four devices: sn0, sn1, sn2, and sn3. Each device can be administered independently.

Besides the sn# interfaces, the multi-link device driver also provides an aggregated IP communication interface, named ml0 (similar to Etherchannel). Only one multi-link can be configured per operating system instance. The ml0 interface distributes all its network traffic over the sn# Switch Network Interfaces. The ml0 IP network interface is configured like any other IP network interface, and it functions like any other IP interface.

Unlike Etherchannel, where the aggregated device takes control and does not allow the use of the component devices, the ml0 allows the independent use of the subsequent sn# interfaces.

The HPS sn# interfaces are designed with redundancy; therefore, the primary benefit of the ml# devices is for performance by bandwidth aggregation. The sn# interfaces automatically handle single-cable faults transparently to the application. Due to this, the HPS requires both protocol devices within a link pair to be sequentially numbered on the same IP subnet. If a second plane is desired, a separate link pair must be assigned.

This section details the steps involved in the installation and configuration of the SNI adapters. Use this procedure when you add the SNI adapters to an existing LPAR. We describe two procedures: One is based on NIM customization, and the second is based on AIX standard installation and configuration.

## Configuring SNIs with NIM

In this section, we describe a general procedure of installing the SNI adapters using NIM. For more details about NIM, refer to *AIX 5L V5.1 Network Installation Management Guide and Reference*, SC23-4385.

To configure SNIs with NIM:

1. Update the NIM master. Refer also to “Configuring the NIM environment” on page 31. Follow these steps to update your NIM server:
   a. Place the required APARs and HPS filesets in the previously defined lpp_source directory using the **gencopy** command. For a list of SNI filesets to copy, see Example 6 on page 30.
   b. Create a new SPOT containing the new APARs and HPS filesets (or update an existing SPOT).
2. Create a stanza file for the SNIs.

   Create a stanza file, which includes the configuration information for the HPS. Each stanza begins with the name of the node and is followed by a series of lines in an “attribute=value” format. The stanza file must contain a stanza for each SNI device to be configured. For further details about stanza file syntax, see *AIX 5L V5.1 Network Installation Management Guide and Reference*, SC23-4385.

   Gather the adapter information using the **getadapters** command:

   ```
   getadapters -z mystanzafile -n node_name
   ```
3. Create a NIM adapter_def resource using the following command:

   ```
   nim -o define -t adapter_def -a server=master -a location=adapter_def_location \
   my_adapter_res
   ```
4. Run the **nimadapters** command:

   ```
   nimadapters -d -f mystanzafile my_adapter_res
   ```

5. Allocate the adapter_def resource to the node:

   ```
   nim -o allocate -a adapter_def=my_adapter_res node_name
   ```

6. At this step, you can install the nodes from scratch (as in Example 14 on page 33) so that the node will have the SNI adapters configured in AIX after the installation.

   You can take into consideration a customization of a node in order to configure the SNI adapters. For example, after the software installation of AIX and the device drivers, you can perform the SNI adapter configuration:

   ```
   nim -o cust -a adapter_def=my_adapter_res node_name
   ```

   After the customization, reboot the server for the changes to take effect.

## Configuring SNIs using standard AIX methods

Attributes of the SNI can be changed using the standard AIX methods.

### *Installing the SNI drivers*

You can use the installp method and the SMIT interface to install the filesets in AIX. The following scenario assumes that the SNI device drivers are downloaded locally, but they can be placed on an NFS file system, too.

Perform the following operations on the LPAR:

1. Create a directory for storing the SNI filesets:

   ```
   mkdir /tmp/SNI
   ```

2. Copy the SNI filesets to the /tmp/SNI directory. Apply the latest patches from IBM® Support Web site.

3. Create the toc file if not created:

   ```
   cd /tmp/SNI
   inutoc .
   ```

4. Run:

   ```
   installp -aX -d '.' all
   ```

5. Reboot the LPAR.

6. Verify that the devices were detected by the system. Example 17 shows an LPAR with one link pair attached.

*Example 17 Listing the SNI devices in AIX*

```
LPAR2# lsdev -C |grep sn
sn0          Defined                Switch Network Interface
sn1          Defined                Switch Network Interface
sni0         Available              Switch Network Interface Adapter
sni1         Available              Switch Network Interface Adapter

LPAR2# lsdev -C |grep ml
ml0          Defined                Multilink Network Interface
mlt0         Available              Multilink Communication Adapter
```

Note that the sn# and ml0 interfaces are in a Defined state, because they are not yet configured in the ODM.

### *Configuring the interfaces*

Configuring the TCP/IP interfaces for the SNI is similar to the configuration of TCP/IP over other interfaces (such as en#). You can either use SMIT or the command line interface:

- Using SMIT menus:
  - Use **smit chsn** for the sn# interfaces.
  - Use **smit chml** for the ml0 interface.
- Using the **chdev** command, the same operation can be done from the command line interface. Example 18 uses a configuration with one link pair in the LPAR.

*Example 18 Assign IP address to sn and ml interfaces*

```
LPAR1# chdev -l sn0 -a netaddr=20.20.20.11 -a netmask=255.255.255.0 -a state=up
sn0 changed
LPAR1#chdev -l sn1 -a netaddr=30.30.30.11 -a netmask=255.255.255.0 -a state=up
sn1 changed
LPAR1# chdev -l ml0 -a netaddr=10.10.10.11 -a netmask=255.255.255.0 -a state=up
ml0 changed
```

List the attributes of the defined interfaces. See Example 19.

*Example 19 List the ODM attributes of the sn and ml interfaces*

```
LPAR1# lsattr -El sn0
mtu     65504          Interface MTU             True
netaddr 20.20.20.11    Network Address           True
netmask 255.255.255.0 Network Mask              True
state   up             Current Interface Status True
LPAR1# lsattr -El sn1
mtu     65504          Interface MTU             True
netaddr 30.30.30.11    Network Address           True
netmask 255.255.255.0 Network Mask              True
state   up             Current Interface Status True
LPAR1# lsattr -El ml0
netaddr 10.10.10.11    N/A True
netmask 255.255.255.0 N/A True
state   up             N/A True
```

- Using the **ifconfig** command.

  Refer to the AIX man page for information about using the **ifconfig** command.

**Note:** In order to configure the ml0 interface, you must configure at least one sn# interface. Regardless of the available number of SNIs on your LPAR, you have a single instance of multi-link device, ml0.

### *Virtual memory manager tuning (VMO)*

It is recommended that you change some VMM parameters when using the sn# or ml0 interfaces. In our test environment, we used the following values:

- Large page size = 16777216
- Large page regions = 64

Use the following command:

```
vmo -r -o lgpg_size=16777216 -o lgpg_regions=64
```

**Note:** You need to reboot for the changes to take effect.

## Verifying the switch connections

To verify the switch connections:

- TCP/IP:
  - Use the **ping** command over the sn and ml interfaces.
  - Use a standard TCP/IP application (Telnet, FTP, rlogin, and so on).
- Check the status in the SNM views on the HMC. See “Administration of the pSeries High Performance Switch” on page 37.
- If you have specific jobs, such as PE jobs or TSM backups, be sure to test this. You might need additional tuning to use the performance capabilities of the HPS fully.

# Administration of the pSeries High Performance Switch

The SNM tasks provide all of the basic switch administration functions, and many of the other functions are handled automatically. As such, none of the CSS E-commands exist for the HPS. Most of the functions are for viewing, but not editing. Command line management can be done through an ASCII terminal interface as the hscroot user.

All of the functions from the SNM GUI are also available from the command line interface. We present these commands in the following sections.

## Power commands

Power commands include:

- **chswnm** - This command enables and disables the SNM. Refer to “Enabling and disabling SNM functions on the HMC” on page 26.
- **chswpower** - This command performs power on/off operations on the switch. You can do the same thing from the GUI interface, as shown in Figure 17.

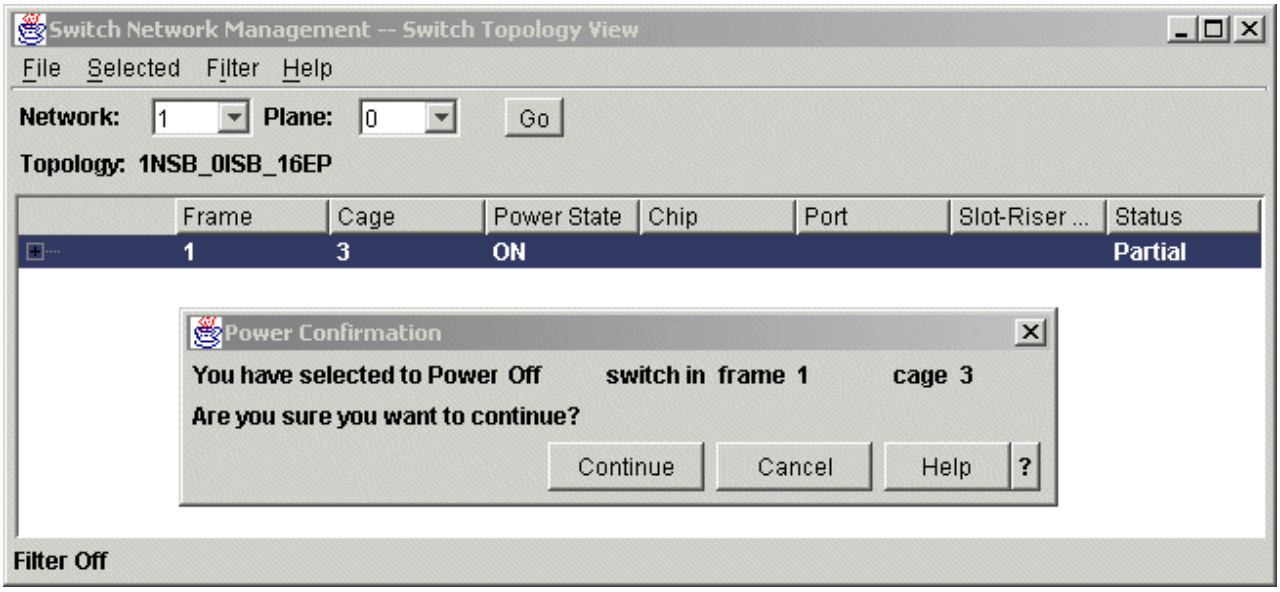


*Figure 17 Switch Network Management: Switch power off screen*

## Diagnostic commands

Diagnostic commands include:

- **`verifylink`** - This command is for concurrent verification of a link.
- **`testlinecont`** - This command is for testing link continuity.

You can access the diagnostic functions over the switch by using the HMC GUI, as shown in Figure 18.

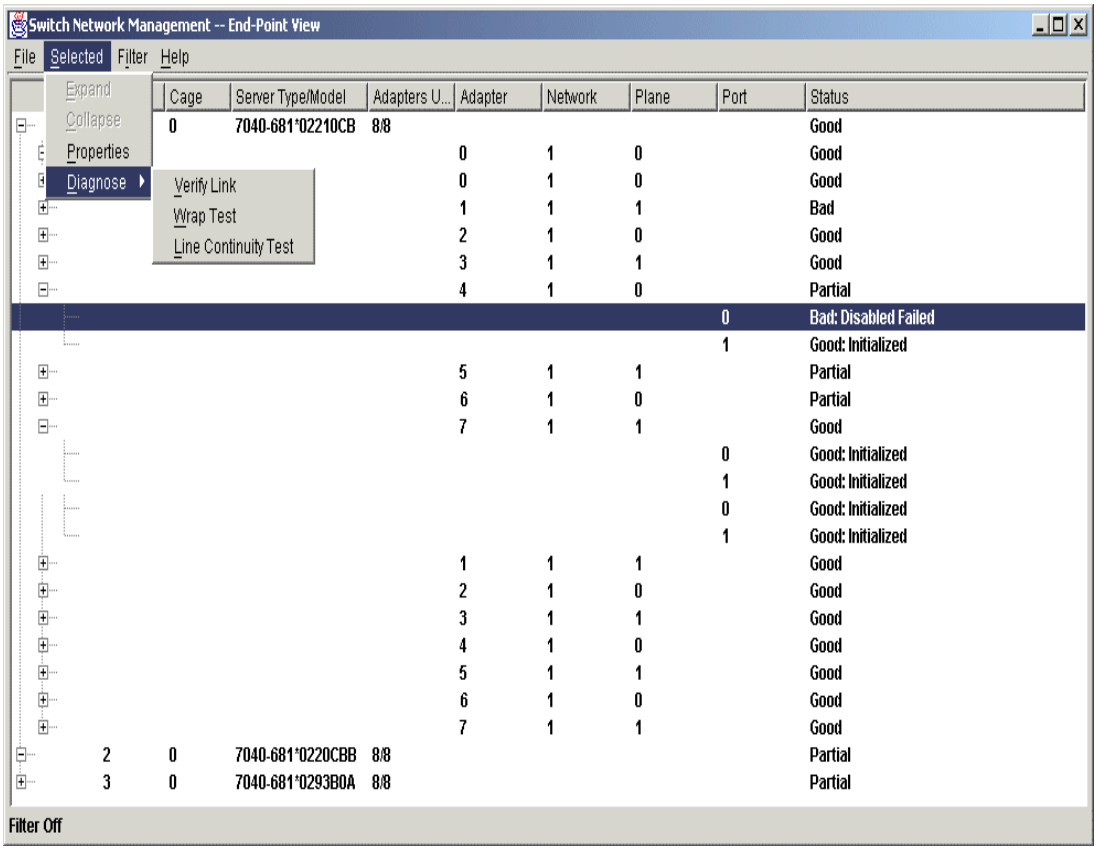


*Figure 18 Switch diagnostic menus*

## Query commands

Query commands include:

- **`lsswendpt`** - This command displays the End-Point View data.

  This provides the same output as End-Point View in the GUI. See Figure 19 on page 39.

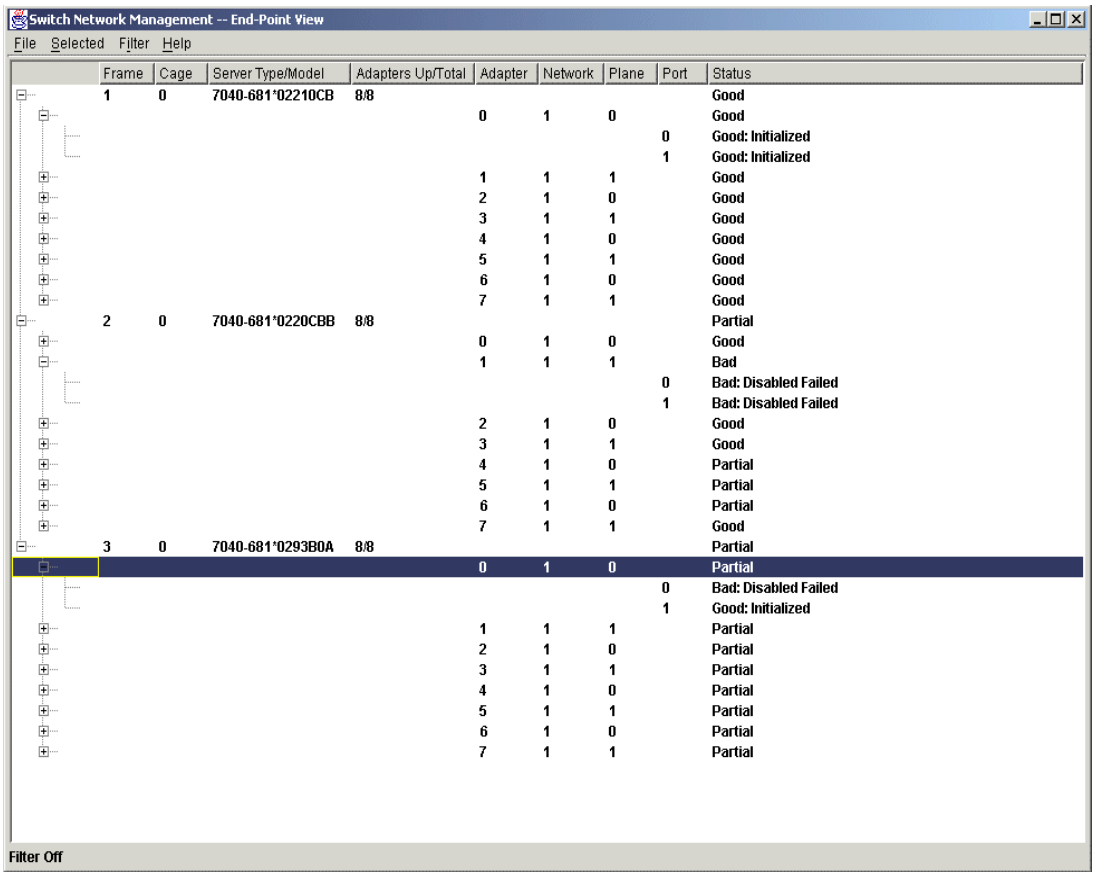


*Figure 19 Switch Network Management: End-Point View data*

- **`lsswtopol`** - This command displays the Topology View data.

  This command provides the same output as the Topology View in the GUI, as shown in Figure 20 on page 40.

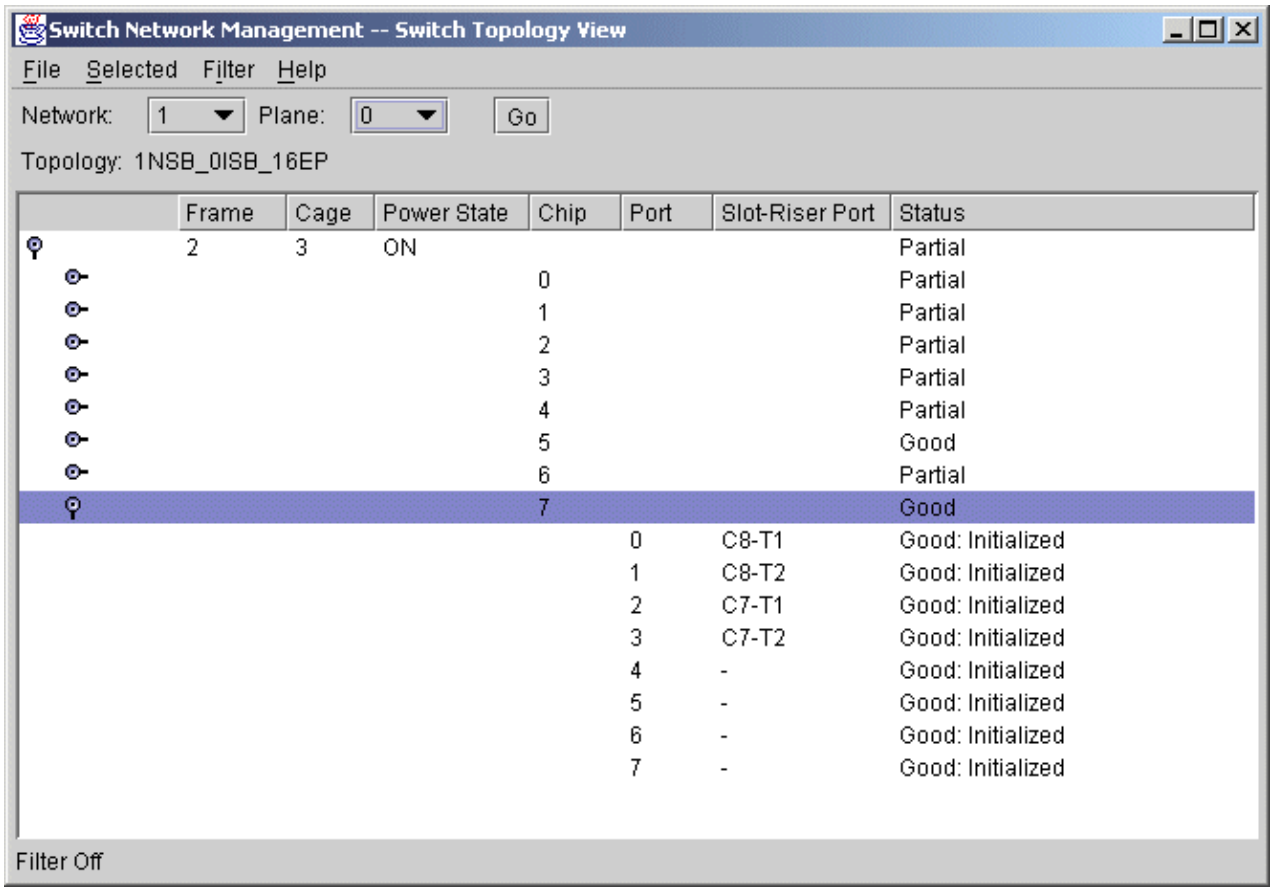


*Figure 20 Switch Network Management: Switch Topology View data*

Note that while this shows the link status, there is no concept of host or switch responds. Link status is purely a hardware function and not a communication function. If your interface is down to AIX, your link might still be good.

- **`lsswenvir`** - This command displays the switch environmentals. The switch power enviromentals information, such as the voltage status and temperature, are gathered from the DCAs. The output looks similar to that shown in Figure 21.

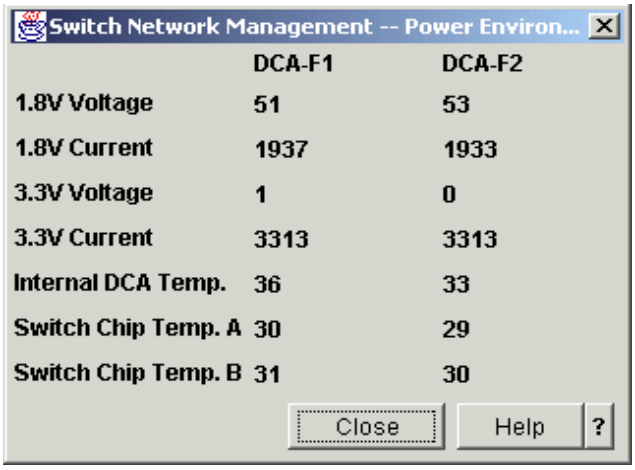


*Figure 21 Display power enviromentals*

- **lsswmanprop** - This command displays the SNM properties, as shown in Figure 22, Figure 23, and Figure 24 on page 42.

  Figure 22 shows the Management view.

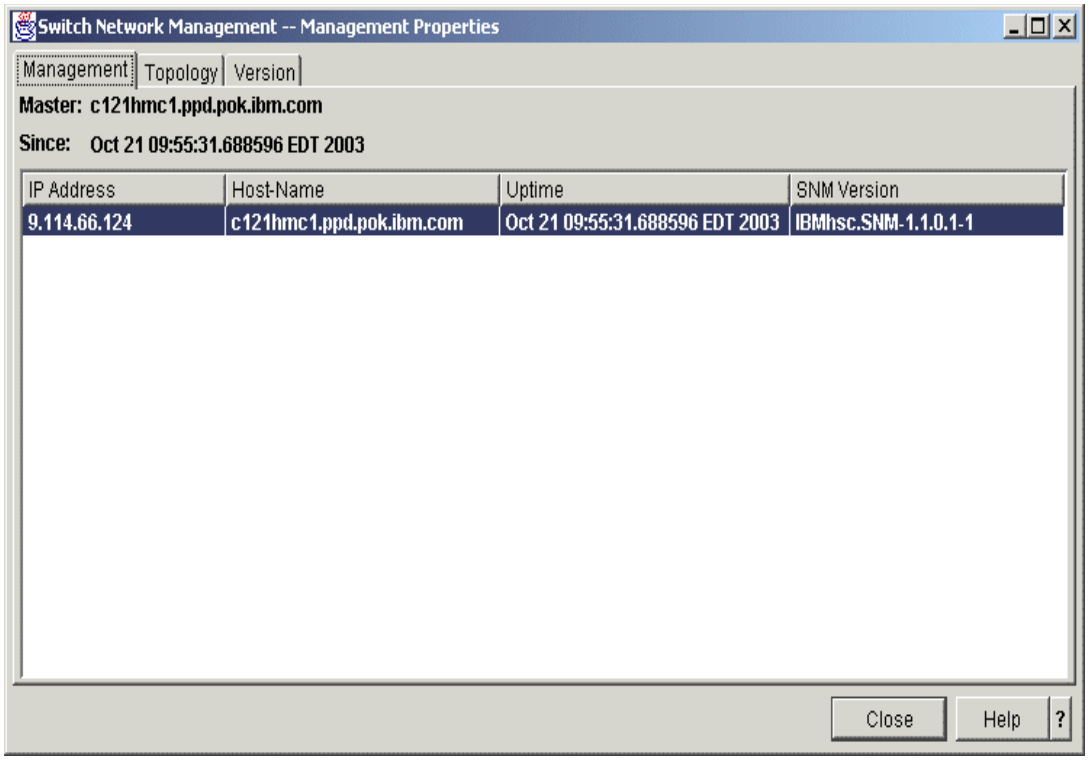


*Figure 22 Switch Management Proprieties: Management view*

Figure 23 shows the Topology view.

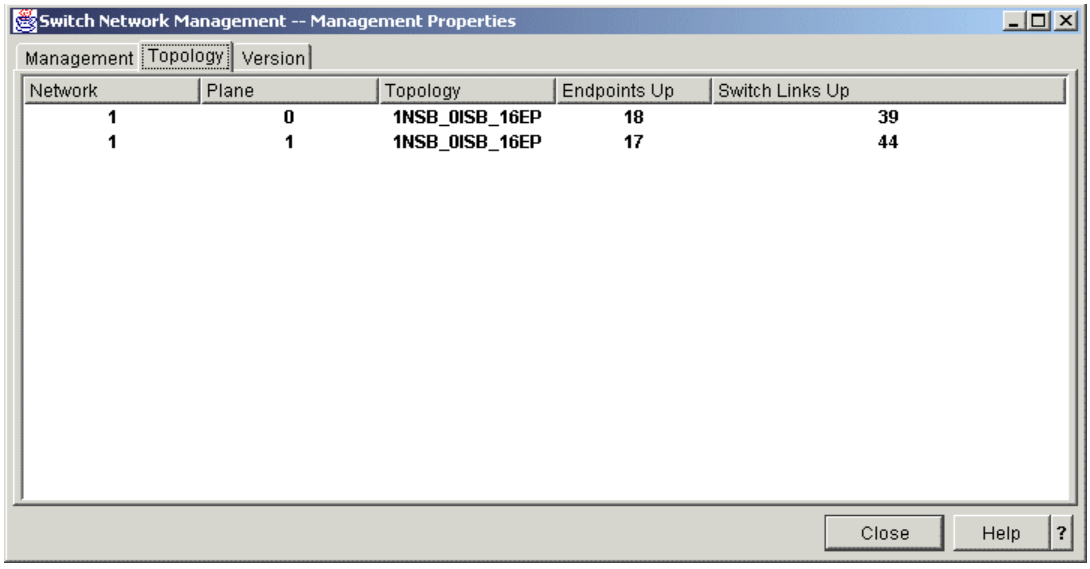


*Figure 23 Switch Management Proprieties: Topology view*

Figure 24 on page 42 shows the Version view.

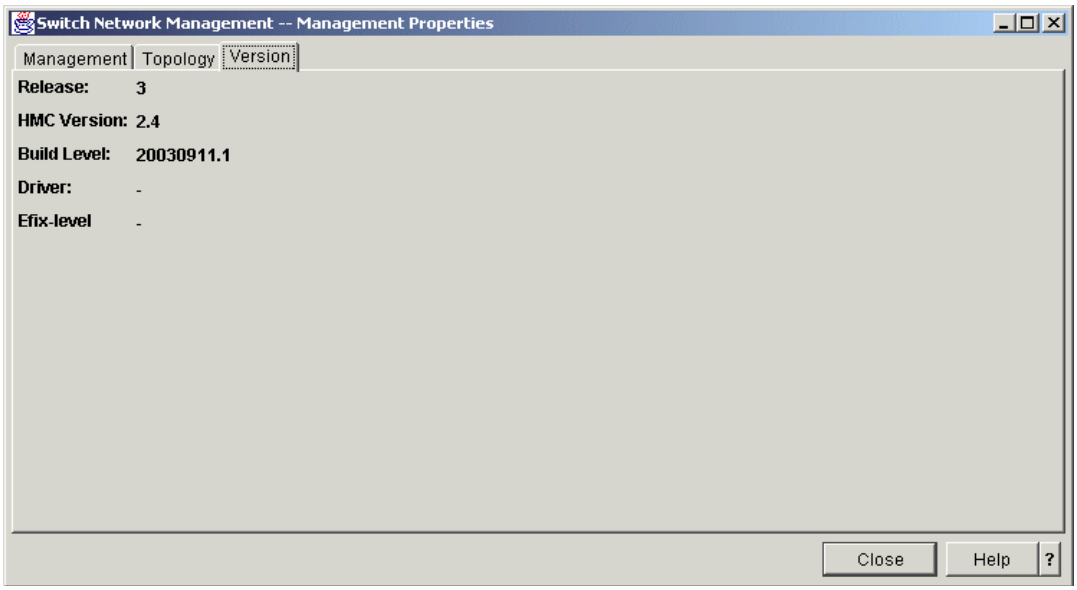


*Figure 24 Switch Management Proprieties: Version View*

- **`lsswtrace`** - This command enables you to view SNM trace log, similar to the GUI screen shown in Figure 25.

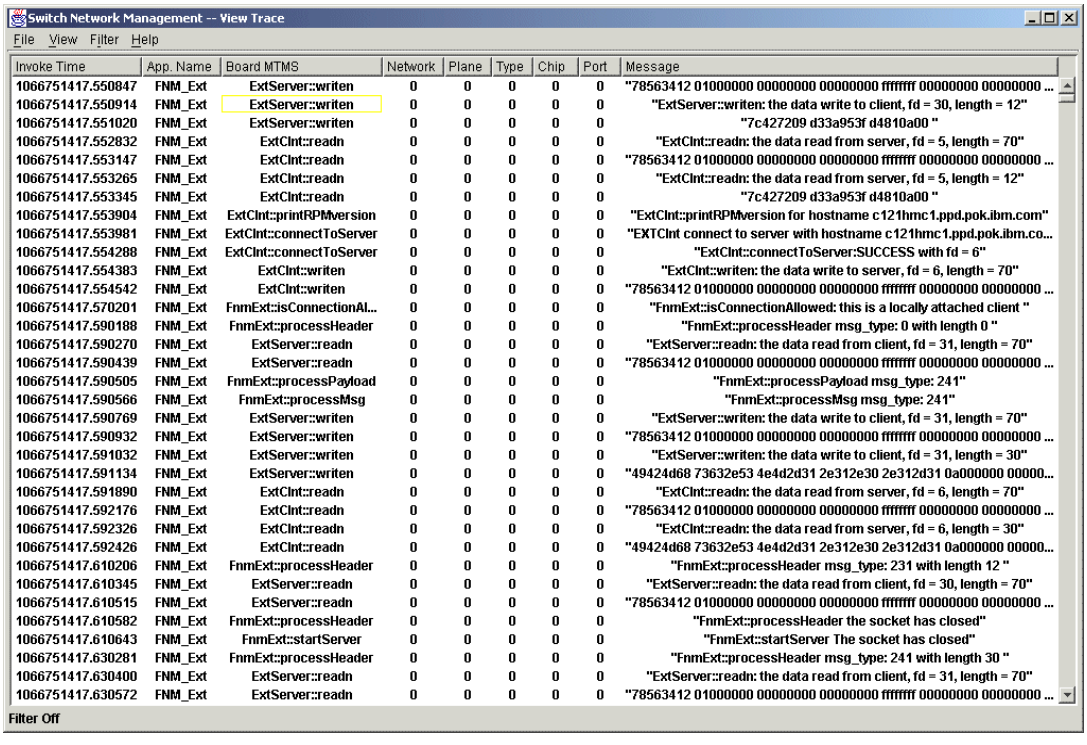


*Figure 25 Switch Network Management: Trace view*

## Switch tuning

The following parameters of the SNI can be tuned for performance purposes. You can change them using the **`chgsni`** command. For further details related to the **`chgsni`** command

and the recommended values, consult *IBM Switch Network Interface for eServer pSeries High Performance Switch Guide and Reference*, SC23-4869.

For SNI, the following ODM attributes specify bounds on window sizes and memory usage:

| | |
|---|---|
| **win_poolsize** | Total pinned node memory available for all user-space receive FIFOs |
| **win_maxsize** | Maximum memory used per window |
| **win_minsize** | Minimum memory used per window |

**Note:** You do not need to reboot your machine after changing these attributes.

# The team that wrote this Redpaper

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

**Octavian Lascu** is a Project Leader at the International Technical Support Organization, Poughkeepsie Center. He writes extensively and teaches IBM classes worldwide on all areas of pSeries clusters and Linux. Before joining the ITSO two years ago, Octavian worked in IBM Global Services, Romania, as SW and HW Services Manager. He holds a master's degree in Electronic Engineering from Polytechnics Institute in Bucharest and is also an IBM Certified Advanced Technical Expert in AIX, PSSP, and HACMP. He has worked with IBM since 1992.

**Pablo Pereira** is an IT Specialist for IBM Global Services in Uruguay serving government and telecommunications customers, responsible for design and implementation of complex solutions. He has five years of experience in the UNIX® and pSeries fields. His areas of expertise include AIX, PSSP, HACMP, and TCP/IP.

**Fernando Pizzano** is a System Administrator in IBM UNIX Development Lab, Poughkeepsie, New York. He has seven years of information technology experience. The last five of those years have been with IBM. His areas of expertise include AIX and IBM @server pSeries hardware. He holds an IBM certification in pSeries AIX 5L System Support. His current position is in the Communication Protocols and Application Tools Development department.

**Zbigniew Borgosz** is a Senior Systems Consultant working for Computerland S.A., an IBM Business Partner in Poland. He joined the company in 1998 and has five years of experience in the UNIX field. His areas of expertise include designing and implementing highly available and scalable UNIX-based solutions (on pSeries, Sun Microsystems, and Hewlett-Packard platforms). He has written extensively about PSSP implementation and problem determination.

**Josh-Daniel Davis** is a Staff Software Engineer for IBM Global Services in Dallas, Texas. He has nine years of experience in information technology and has been with IBM for more than five years. His areas of expertise include AIX, Linux, and pSeries servers. He is certified for Tivoli® Storage Manager and is a Certified Advanced Technical Expert for AIX 4.3 and 5L, including PSSP and p690 support.

**Andrei Socoliuc** is a Software Support Engineer with IBM Global Services in Romania. He holds a master's degree in Computer Science from Polytechnic Institute in Bucharest, Romania. He has six years of experience in the pSeries clusters field. His areas of expertise include AIX, PSSP, HACMP, TSM, and Linux. He has written extensively about pSeries clusters managed by PSSP.

# 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 illustrates 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. 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 IBM's application programming interfaces.

Send us your comments in one of the following ways:

- Use the online **Contact us** review redbook form found at:
  **ibm.com**/redbooks
- Send your comments in an Internet note to:
  redbook@us.ibm.com
- Mail your comments to:
  IBM Corporation, International Technical Support Organization
  Dept. JN9B  Mail Station P099
  2455 South Road
  Poughkeepsie, NY 12601-5400 U.S.A.

# Trademarks

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

| | | |
|---|---|---|
| AIX® | IBM® | Redbooks(logo) ™ |
| AIX 5L™ | ibm.com® | Tivoli® |
| @server™ | POWER4™ | xSeries® |
| @server™ | POWER4+™ | |
| eServer™ | pSeries® | |

The following terms are trademarks of other companies:

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

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

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

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