Introduction

IBM pSeries High Performance Switch (HPS) Service Pack 21 supports new GFW and AIX levels, as well as new HPSNM code levels.

This Service Pack details updates to these components:

- Hardware Management Console (HMC) 3.3.7 - Build level 20060207.1
- Global Firmware (GFW) 3H061030/3J061030
- Power Subsystem Microcode (PT Code) 26A9
- SwitchNetwork Manager (SNM) 1.3.7.0
- HPS/SNI Devices LP Software (AIX 5.3) 1.2.0.6
- HPS/SNI Devices LP Software (AIX 5.2) 1.1.3.10
- AIX 5L Version 5.3 5.3.0.54/5300-05-04
- AIX 5L Version 5.2 5.2.0.98/5200-09-03

AIX High Performance Computing (HPC) and Clusters LPs as listed in the Detailed LP Level Check

- VSD
- LAPI
- HPS/SNI Devices
- PPE
- LoadLeveler
- GPFS
- CSM
- RSCT

This Service Pack also contains general guidelines for upgrading the components listed in the **Component updates/download information** section.

These guidelines are intended to be a supplement to the other IBM documents referred to in this document. We strongly advise that you have the referenced documents available before you begin the

upgrade process.

A list of referenced documents can be found in the **Supporting Documentation** section. The Code Levels listed in the **Component updates/download information** section reflect the levels available at the time of this Service Pack release.

Some components support only a single version, notably the Microcode for GFW and the Power Subsystem. Subsequent released versions are expected to be backward compatible.

The procedure outlined in **Installation Sequence** section is the standard sequence of installation. Non-standard sequences or undocumented code levels may cause unforseen problems. In this event please contact your Customer Service Representative.

If you are upgrading to this Service Pack from a service pack that is earlier than SP9, then installing CSM 1.4 is required. The recommended AIX service level for AIX 5L version 5.2 is TL 08 SP01. The recommended AIX service level for AIX 5L version 5.3 is TL 05.

The Maintenance package contains CSM 1.4 which requires RPM update openCIMOM 0.8(5.2).

Because '/var' is a system data repository, system administrators should check periodically to maintain /var such that there is at least 30 Mb free (use the **df -k** command). If it is more than 75% full, look for the directories that contain the most data (use the **du /var | sort -n** command).

Supporting documentation

Refer for to the following documentation for additional information to supplement the HPS Readme.

- Hardware Management Console
- Hardware Management Console for pSeries Installation and Operations Guide
- Hardware Management Console for pSeries Maintenance Guide (SA38-0603-05)
- pSeries High Performance Switch (HPS) Planning, Installation and Service Guide (GA22-7951-01) (HPS Guide)
- IBM eServer pSeries 690
- Switch Network Interface for eServer pSeries High Performance Switch Guide and Reference (SC23-4869-01)
- AIX 5.2 documentation
- AIX 5L Version 5.2 Installation Guide and Reference
- AIX 5L Version 5.2 Performance Management Guide
- Reliable Scalable Cluster Technology (RSCT) Library
- General Parallel File System (GPFS) Library
- Cluster System Management (CSM) Library
- IBM Parallel Environment for AIX 5L Installation Version 4 Release 1.1
- IBM LoadLeveler for AIX 5L and Linux Using and Administering Version 3 Release2
- LoadLeveler 3.2 documentation updates
- pSeries and AIX Information Center
- AIX 5L Version 5.3 documentation
- System management guides
- Installation guides

Software updates

The following software updates are included:

- New SNM/FNM features
- MPI and LAPI performance enhancements
- Support for Node Switch Board (NSB) and Intermediate Switch Board (ISB) failure
- Application Striping and Fail-Over of packets across multiple interfaces for fault resilience
- Application Checkpoint/Restart and preemption
- The Web-based System Manager Remote Client (WebSM PC CLIENT may need to be reinstalled
- RDMA (remote direct-memory access) transport on HPS

New SNM features

This section discusses the new features and functions on the Switch Network Management Panel introduced in HMC 1.3.1.0 Service Pack 9.

For details of operation please see Chapter 4. System management components and Appendix E. Switch Network Manager (SNM) in pSeries High Performance Switch (HPS) Planning, Installation and Service Guide (GA22-7951-02) (HPS Guide)

Features

HMC Fail-over:

- FNM/SNM daemon Enable SNM Software in Normal Mode runs on up to four HMCs in the cluster.
- Verification mode: Enable SNM Software for Switch Network Verification used to check out the cluster.

New features:

- Select Logical Topology
- Display Cluster Components
- Enable SNM Software for Switch Network Verification

The "Enable SNM Software for Normal Mode" has been renamed. It used to be "Enable SNM Software"

Select Logical Topology introduces a requirement to set the logical topology of your system before enabling SNM software. If the logical topology has not been set, neither of the tasks to Enable SNM Software will be selectable. This is required. See note 1.

Note: If you are using the SNM Fail-over support introduced in the Software Update release, the logical topology must be set on each HMC where you enable SNM software.

Display Cluster Components shows the frames and cages of the cluster components with which SNM can communicate. This task can only be used if SNM software is *not* enabled. The lsswcomp command provides the same function as the "Display Cluster Components" task.

Enable SNM Software for Switch Network Verification task is used to check out the cluster by NOT removing bad links or routes. The main purpose for this feature is to discover cable mis-wires and defective hardware. No Routing is modified or loaded. It is NOT intended to be used in a working environment.

Enable SNM Software for Normal Mode. Select this feature to run SNM in a normal, working environment. This is required. See note 1.

Notes

Notes:

1. If you are upgrading from Service Pack 8 or less, these steps are *required* for an initial installation of FNM/SNM:

   First Select Logical Topology, and then select Enable SNM Software for Normal Mode. See "Enable SNM Software" in the HPS Guide.
2. The systems topology selected needs the number of NSB' and ISB's calculated on a per plane basis.

   Example: For a configuration with two NSB's and two planes, there is one NSB per plane with 16 endpoints per plane. This equates to 1 Network, 2 Planes, and 1NSB_OISB_16EP.

LAPI Performance Enhancements
LAPI and MPI host communication stacks have been tuned, based on experience, for lower latency. These changes complement the communication performance improvements released in part 1. They are being released as one package with the LAPI and MPI striping function.

Note: Beginning with the Service Pack 9:

- A single MPI user-space job which wants to use HPS adapters in both of two switch planes must set **MP_EUIDEVICE=sn_all**(or csss). In previous releases, a single MPI/us job with multiple tasks per node could use adapters from two planes by setting MP_EUIDEVICE=sn_single (or css0).
- A job which sets MP_EUIDEVICE=sn_single will only be able to use half of the HPS adapters on the nodes where it runs.
- A job must set MP_EUIDEVICE=sn_all in order to use all the HPS adapters in the nodes it is using.

Support for Node Switch Board (NSB) and Intermediate Switch Board (ISB)
The SNM daemon has been enhanced to ensure that the failure of an NSB and an ISB will be handled appropriately. The main enhancement is that the Switch Network Manager daemon gathers and downloads multiple path table updates into a single transaction instead of one update per transaction. An entire switch board failure or recovery will generate multiple of path entries changes. Therefore handling these updates in groups involves many fewer transactions. These can be processed in a more timely fashion and place a smaller load on SNM and the service network.

Application striping and fail-over of packets across multiple interfaces for fault resilience
Striping provides a method for a single task of a parallel application to utilize multiple SNI links. This allows a single link to fail without the application using the link being terminated because there is an alternate path to all other tasks of the application. The striping method implemented is designed to provide resilience to switch or interface/link failure and is not designed to provide an increase in the aggregate network bandwidth as seen by tasks.

This link failure resilience function requires support in LoadLeveler,Parallel Environment and the host communication stack for MPI and LAPI. When an interface/link recovers, the striping function will ensure that the link is used again. Another goal of this striping design is to evenly distribute traffic over multiple switch networks to maintain a balance in the communications load.

Known issues and concerns

Known issues and concerns include the following:

- In non-striping mode you only get half the links on a p655 system using a 2 plane configuration when setting MP_EUIDEVICE=sn_single. You must use MP_EUDEVICE= csss or sn_all to get all links.
- Performance degradation for single task per link of 3% for unidirectional and 5% for exchange bandwidth at large messages. However, multiple tasks per link get full link bandwidth.
- Striping currently limited to two links per task. Multiple tasks can use all links as long as there are more tasks than half the links available.

Note: When you are using RDMA, there is no large message performance penalty for striping. In fact striping greatly improves large message RDMA bandwidth. The above performance degradation occurs only in packet mode.

Application Checkpoint/Restart and Preemption
Checkpoint/restart provides a way to stop and resume applications at some later time. It is a very useful tool in managing the clusters workload since applications can be check-pointed to allow other applications to run or the system made inactive for a maintenance window.

Preemption is also provided so a running application can be suspended to allow another higher priority application to execute right away.

The Web-based System Manager Remote Client (WebSM PC CLIENT)
If the Web SM is updated, it is recommended that the existing code be reinstalled on your remote server or PC to ensure that the WebSM is compatible with this HMC Build.

To install the PC Client software on your remote server or PC go to: http://<hmc-hostname>/remote_client.html.

The two versions are the "legacy" WebSM client - "InstallShield" and "The future of WebSM clients." - Webstart. Both versions can exist simultaneously on your client workstation
Install Shield

If there is a currently installed install shield version of WebSM, it *must* be **uninstalled** before the new version is installed.

To uninstall, follow these steps:

1. From the Start Menu, select Control Panel -> Add/Remove Programs.
2. Select Web-based System Manager Remote Client, and then click "Change/Remove".
3. Follow the on screen instructions.

Failure to uninstall will result in undefined behavior of the WebSM PC Client.

To install, follow these steps:

1. Go to http://<hmc-hostname>/remote_client.html.
2. Select "Install Shield".
3. Select the Operating System where the program will be running.
4. Select "Open" to immediately install on the machine on which the browser is running.
5. Selecting "Save" allows you to store the install program for installation at a later time.

Note: The Install Shield package is a 100 Mb file, so either process will require about 15 minutes. Once saved, the self-extracting file can be used to upgrade multiple PCs.

Webstart

The webstart installation is a two part process:

1. Installing the 1.4.2 JVM that contains the webstart code.
2. Invoking the launch code and gets the initial classes downloaded from the HMC.

When the client is launched, it checks for new jar files on the HMC and downloads them, so subsequent updates of the HMC server code do not require the updates of the webstart client code.

Having multiple jvms installed on the client workstation may result in the default jvm path not pointing to the IBM 1.4.2 jvm. Since webstart does not work well with the Sun 1.4.2 jvm, you may need to clean up multiple jvm installations on your workstation.

The Java Web Start version requires multiple logins on launching. You have to log in to the original HMC whenever you try to connect even if you are just downloading the webstart **.jsp** files. Downloading the **.jsp** files from a regular AIX machine requires one less login.

Multiple logins are not an issue in the non-webstart websm because you are not connecting to the webserver to download the .jsp files. Trying to 'cancel out' results in multiple (up to 16) presentations of

the login verification box.

1. Install Java Web Start:
    - Web Start for Linux- Install Java Web Start on a Linux platform. Once installed, return to this page to download the Remote Client.
    - Java Web Start for Windows- Install Java Web Start on a Windows platform. Once installed, return to this page to download the Remote Client.
2. Download the Web-based System Manager Remote Client for Java Web Start on Linux and Windows systems.

RDMA transport on HPS
The IBM eServer pSeries High Performance Switch (HPS) with Remote Direct Memory Access (RDMA) enabled delivers 1.8GB/s of unidirectional bandwidth across a single link and over 3.5GB/s over two links to a single MPI task for some applications. Bidirectional bandwidth measures almost 3.0GB/s over a single link and over 5.9GB/s for a single MPI task striping over two links.

These measurements were achieved on two nodes containing 1.9GHz, Power 4+ based processors with 32 CPUs and two HPS adapters, and links on each of the two nodes. Large pages and cache-line aligned buffers were used for the MPI measurements. These results indicate nearly linear scaling of large message striping performance. A single link is able to deliver over 1.65GB/s of bandwidth over TCP/IP.
Advantages or RDMA
RDMA provides the following advantages:

- Decouples the CPU from the movement of data allowing for better overlap of computation and communication.
- Reduces the computational load of the CPU by off-loading segmentation and reassembly of messages to the network adapter, thereby reducing the number of packet arrival interrupts.
- Reduces the stress on the memory subsystems by reducing the number of bus crossings - one I/O bus crossing versus the traditional two memory bus and one I/O bus crossing when not using RDMA.
- Enables the protocols to efficiently stripe a message (or different messages) from a single task across multiple network interfaces to exploit the available communication bandwidth in parallel without engaging multiple CPUs.
- Provides improved raw transport performance. In cases where the transport bottleneck is the copy rate (memory bandwidth), RDMA helps eliminate that bottleneck.

How application can exploit RDMA

To exploit the overlap potential of RDMA, applications should make use of non-blocking calls and schedule communication as early as possible to extract the full overlap potential. For MPI/LAPI applications the RDMA transport kicks in for large messages (to justify the setup costs of RDMA).

Applications that reuse the same communication buffers will benefit further from RDMA since the setup cost is amortized over larger number of transfers. For applications that send large messages asynchronously, enabling striping should help the applications realize superior transport bandwidth.

RDMA transport is **disabled** by default in this code release, to insure that customers running production workloads that do not need RDMA operate without change.

Note:
When RDMA is enabled by changing the device attribute **rdma_xlat_limit** to a value greater than 0, it is possible for an RDMA enabled job to surpass the AIX pin limit which may lead to a hung system. If this pin limit is exceeded slowly enough the parallel application will catch a SIGDANGER signal and kill the job. However, if this pin limit is exceeded too rapidly, AIX may resort to killing processes, including systems processes, which may cause system to appear to be hung. This situation may be prevented by setting the rdma_xlat_limit to a number which is smaller than 4096 times the sum of the number of large and small pages on the system.

RDMA usage

By default, the RDMA capability of the SNI adapters is turned off. Use the chdev (or optionally /usr/sni/aix52/chgsni) command to turn on RDMA by setting rdma_xlat_limit to a value greater than 0x0.

The rdma_xlat_limit attribute controls amount of memory allowed for RDMA, per LPAR. The default value of the rdma_xlat_limit attribute is 0x0, effectively turning RDMA off.

When using RDMA it is recommended that this attribute value be set to value equal to 75% of small page memory. This will allow all of the large pages on the system and most of the small pages to be mapped to the sni adapters.

Note:
This odm attr (rdma_xlat_limit) is changed for ALL sni's in an LPAR with just the single command. For the change to take effect, the LPAR must be rebooted. The RDMA option MUST be turned ON OR OFF across the entire cluster. If it is not, SNI adapters will not communicate due to version mismatch failures.

When RDMA is turned on, the IP protocol automatically uses the RDMA capability. For user space jobs, an additional LoadLeveler keyword (bulkxfer=yes) must be set in the LoadLeveler job control file to indicate that this job is requesting RDMA. This allows users to run both RDMA and non-RDMA jobs on the system.

For the most current details, refer to LoadLeveler 3.2 documentation updates; Addition of support for Bulk Data Transfer.

For additional information, refer to LoadLeveler Using and Administration Guide.

To enable RDMA:
/usr/sbin/chdev -l sniX -a rdma_xlat_limit=XLAT_LIMIT

where XLAT_LIMIT is 75% or less of the small page memory.

To disable RDMA:
/usr/sbin/chdev -l sniX -a rdma_xlat_limit=0x0000000000000000

To check the RDMA status: -- run lsattr -El sni0 ; lsattr -El sni1 ;

Example:

for i in `lsdev | grep sni.*Avail | awk '{print $1}'`; do lsattr -El $i|grep rdma; done

chdev Usage: chdev -l Name [-a Attribute=Value]...[-p ParentName][-P|-T]
chdev -l Name -a rdma_xlat_limit=size

The 'True' seen in the lsattr output is whether or not the value is user modifiable. The rdma_xlat_limit attribute is one of 5 that are not valid on just 1 sni. The change method updates the odm for all snis.

This attribute is modified for all existing devices. Any user supplied -l option is ignored.

Attribute values for rdma_xlat_limit :
Default value: 0x0 (Implies RDMA is turned off)
MINimum value: 0x0
MAXimum value: 0xFFFFFFFFFFFFFFFF (Maximum amount of memory allowed for RDMA (in bytes))
Recommended value: 75% or less of small page memory.
Regarding the use of memory by RDMA and TLP

Real memory is divided into two categories - Small Pages and Large pages. It is the user's responsibility to achieve an optimal balance between the the two categories based on the expected and/or experienced needs of both SNI adapters memory requirements expressed in TLP and applications use of Small Pages as expressed in RDMA. TLP can allocate up to 75% of real memory. RDMA can pin and map up to 75% of small page application memory.

Total Real Memory is a function of N(bytes of real mem) = T(bytes of reaL mem allocated to TLP) + S (bytes of real mem allocated to Small Pages).

Small Page memory is a function of S(bytes of real mem allocated to Small Pages) = N(bytes of real mem) - T(bytes of reaL mem allocated to TLP).

Large Page memory is a function of T(bytes of reaL mem allocated to TLP) = N(bytes of real mem) - S (bytes of real mem allocated to Small Pages).

The amount of small page memory can be calculated as follows:

lsattr -E -l sys0 -a realmem returns the number of kbytes real memory, call this number A.
vmo -o lgpg_regions returns the number of large pages, call this number B.
Then A*1024 - B*16*1024*1024 is the amount of small page memory in bytes.

Example:

```
#!/bin/ksh
real_mem=`lsattr -E -l sys0 -a realmem|awk '{print $2}'`
lgpg_regions=`vmo -o lgpg_regions|awk '{print $3}'`
A=$(( real_mem * 1024 ))
B=$(( lgpg_regions * 16*1024*1024 ))
print "Real Mem=$A, TLP=$B; Small pages=$((A - B))"
```

Real Mem=32212254720, TLP=4294967296; Small pages=27917287424

The rdma_xlat_limit will limit the amount of memory that a user application can pin and map for use with RDMA. This pinning and mapping only survives as long as the job it executing. After it exits the memory is unpinned and freed.

See also Regarding the use of memory by TLP.

Component updates

This section lists the service packs for the various components and provides information or links for downloading these service packs. It also details which releases of AIX LPs (components) are compatible with which AIX release and provides detailed level checks for the AIX LPs (components).

- Service pack and download information
- AIX Licensed Products

Service pack and download information

The following table lists service pack information for each component and provides links for downloading or ordering the service packs. In addition, the Component name link displays installation information for the component.

Component	Service pack updates	Download sites
HMC	The HMC code level for this Service Pack is HMC V3R3.7 Choose HMC_Update_V3R3.7.zip APAR IY82939 PTF U807279 This Service Pack does updates this component. **Note:** CD images cannot be downloaded directly from this site. Contact your IBM Sales Representative or Business Partner, and order Hardware Feature Code (MES) 0960 for the initial upgrade CDs. **Note:** If you plan to use the removeable media, please read the Important Notice regarding the use of the Format Removable Media task: HMC POWER4 updates	HMC corrective service OR HMC POWER4 updates

SNM/FNM	The SNM code level for this Service Pack is SNM Version: 1.3.7.0-1 does not update this component. STATUS RPM Version: 1.3.7.0 SNM Version: IBMhsc.SNM-1.3.7.0-1.i386	HMC corrective service OR HMC POWER4 updates
GFW	The Firmware level for this Service Pack is ROM Level (alterable)..... 3H061030 (P690) ROM Level (alterable)..... 3J061030 (P655) Ucode Version = 0x50b101100000000 built on 11/16/05 at 17:00 This code is machine specific. Use option 1, **Download microcode by machine type and model** 1) Download microcode by machine type and model For p690 models: Select "7040-681", and then choose "Version 3H061030" For p655 models: Select "7039-651", and then choose "Version 3J061030" This Service Pack updates this component. **Note:** GFW is available in IBM CORE 3-4 days earlier than the above mentioned website. Please contact your IBM CE for the GFW in IBM CORE if not available on website. Refer to this website for detailed download and unpacking procedures: Microcode downloads	Microcode downloads

Power Subsystem Microcode	The Power Subsystem Microcode level for this Service Pack is 26a9. Go to "Power Subsystem for 7039-651 (p655) and servers containing the 7045-SW4 (High Performance Switch)" This Service Pack does not update this component. Code is the same for both p690 and p655: ptcode-1.78.26a9-1.i386.rpm **Note:** Power Subsystem Microcode is available in IBM CORE up to one week earlier than the above mentioned website. Please contact your IBM CE for the Power Subsystem Microcode in IBM CORE if not available on website.	Microcode downloads
AIX 5.3	The AIX 5.3 level for this Service Pack is 5.3.0.54 AIX53: 5300-05-04 This Service Pack updates this component.	System p support fixes
AIX 5.2	The AIX 5.2 level for this Service Pack is 5.2.0.98 AIX52: 5200-09-03 This Service Pack updates this component.	System p support fixes
CSM openCIMOM	CSM 1.4 requires RPM update openCIMOM 0.8 (5.2) If you are APPLYING software for csm.server 1.4.1.1, please install the following images from the AIX Installation Media Volume 2: • tcl • tk • expect • conserver-8.1	openCIMOM update

AIX Licensed Products (LP)

The following tables address Licensed Product (LP) components of AIX relevant to High Performance Computing (HPC) and Clusters software in this service pack. You need to know which releases of the LP components are compatible with your AIX release, as well as the latest fileset fix levels available at the time this HPS Service Pack was released.

AIX 5.3 and AIX 5.2 Licensed Products (LPs)

The following table lists LPs for AIX 5.3 and 5.2 and indicates which LP releases are compatible with which AIX version.

LP	AIX 5.3	AIX 5.2
VSD	4.1.0	4.1.0
LAPI	2.4.3 2.4.2	2.3.3
HPS/SNI (Devices)	1.2.0	1.1.3
PPE	4.3.0 4.2.2 4.1.1	4.2.2 4.1.1
LoadLeveler	3.4.0 3.3.2	3.3.2 3.2.0
GPFS	3.1.0 2.3.0	2.3.0 2.2.1 2.1.0
CSM	1.6.0 1.5.1 1.4.1	1.5.1 1.4.1
RSCT	2.4.6	2.3.10
ESSL	4.2.0	4.2.0
Parallel ESSL	3.3.0 3.2.0	3.2.0

Detailed LP level check

The LP fileset fix levels listed in the following table indicate the latest available levels at the time this HPS Service Pack was released. They are available from the following website:

Fix Central: System p support fixes

To order the latest filesets for all LP components of this HPS Service Pack, click the APAR link for your Version of AIX:

AIX 5.3: APAR IY93823

AIX 5.2: APAR IY93823

To order the latest APAR for an individual LP component, find the LP name in the table, and then click the APAR link that corresponds to your Release of AIX. The link opens the Packaging Options page in Fix Central, where you select your OS level, and then Continue to the download page.

Alternatively, clicking on th aforementioned Order ALL Fixes link will open a new page on the 'System p support fixes' website. There you will be presented with an opportunity to get all the latest available fixes for this Service Pack.

The selected APAR package will contain fix levels that are the same or higher than the levels listed below.

To check the LP service levels, on each logical partition issue:
lslpp -Lc | egrep "vsd|LAPI|HPS|sni|ppe|LoadL|mmfs|rsct|csm|essl|pessl" | cut -d : -f 2,3 |sed 's/:/ /'

LP	Release	Component ID	OS Ver./APAR	Level check
VSD	410	5765G2602	AIX 5.3 IY93799 AIX 5.2 IY93799	rsct.vsd.cmds 4.1.0.17 rsct.vsd.rvsd 4.1.0.15 rsct.vsd.vsdd 4.1.0.19 rsct.vsd.vsdrm 4.1.0.7
LAPI	243	5765G2601	AIX 5.3 IY93802	rsct.lapi.nam 2.4.3.0 rsct.lapi.rte 2.4.3.2 rsct.lapi.samp 2.4.3.0
	242	5765G2601	AIX 5.3 IY93801	rsct.lapi.nam 2.4.2.2 rsct.lapi.rte 2.4.2.8 rsct.lapi.samp 2.4.2.0
	233	5765G2601	AIX 5.2 IY93800	rsct.lapi.nam 2.3.3.2 rsct.lapi.rte 3.3.3.7 rsct.lapi.samp 2.3.3.0

HPS/SNI (Devices)	120	5765G2400	AIX 5.3 IY93805	devices.chrp.IBM.HPS.rte 1.2.0.6 devices.common.IBM.sni.ml 1.2.0.2 devices.common.IBM.sni.ntbl 1.2.0.0 devices.common.IBM.sni.rte 1.2.0.7 devices.msg.en_US.chrp.IBM.HPS.rte 1.2.0.2 devices.msg.en_US.common.IBM.sni.ml 1.2.0.0 devices.msg.en_US.common.IBM.sni.ntbl 1.2.0.1
	113	5765G2400	AIX 5.2 IY93804	devices.chrp.IBM.HPS.rte 1.1.3.10 devices.common.IBM.sni.ml 1.1.3.2 devices.common.IBM.sni.ntbl 1.1.3.0 devices.common.IBM.sni.rte 1.1.3.9 devices.msg.en_US.chrp.IBM.HPS.rte 1.1.3.1 devices.msg.en_US.common.IBM.sni.ml 1.1.3.1 devices.msg.en_US.common.IBM.sni.ntbl 1.1.3.1
PPE	430	5765F8300	AIX 5.3 IY93808	ppe.perf 4.3.0.1 ppe.poe 4.3.0.1
	422	5765F8300	AIX 5.3 IY93807 AIX 5.2 IY93807	ppe.man 4.2.2.2 ppe.perf 4.2.2.3 ppe.poe 4.2.2.7 ppe.pvt 4.2.2.0
	411	5765F8300	AIX 5.3 IY93806 AIX 5.2 IY93806	ppe.man 4.1.1.0 ppe.perf 4.1.1.4 ppe.poe 4.1.1.10 ppe.pvt 4.1.1.0
LoadLeveler	340	5765E6900	AIX 5.3 IY93812	LoadL.full 3.4.0.2 LoadL.msg.en_US 3.4.0.1 LoadL.so 3.4.0.2 LoadL.webui 3.4.0.2
	332	5765E6900	AIX 5.3 IY93811 AIX 5.2 IY93811	LoadL.full 3.3.2.8 LoadL.msg.En_US 3.3.2.5 LoadL.msg.en_US 3.3.2.5 LoadL.so 3.3.2.8

	320	5765E6900	AIX 5.2 IY93810	LoadL.full 3.2.0.20 LoadL.msg.En_US 3.2.0.7 LoadL.msg.en_US 3.2.0.7 LoadL.so 3.2.0.19 LoadL.tguides 3.2.0.1
	310	5765F64AP	AIX 5.3 IY93815	gpfs.base 3.1.0.8 gpfs.docs.data 3.1.0.3 gpfs.msg.en_US 3.1.0.7
	230	5765F64AP	AIX 5.3 IY93814 AIX 5.2 IY93814	gpfs.base 2.3.0.18 gpfs.docs.data 2.3.0.8 gpfs.msg.en_US 2.3.0.12
GPFS	221	5765F64AP	AIX 5.2 IY93813	mmfs.base.cmds 3.6.1.10 mmfs.base.rte 3.6.1.13 mmfs.gpfs.rte 2.2.1.12 mmfs.gpfsdocs.data 3.6.1.4 mmfs.msg.en_US 3.6.1.4
	210	5765F64AP	AIX 5.2 IY84773	mmfs.base.cmds 3.5.0.13 mmfs.base.rte 3.5.0.22 mmfs.gpfs.rte 2.1.0.25 mmfs.gpfsdocs.data 3.5.0.5 mmfs.msg.en_US 3.5.0.8
CSM	160	5765E88AP	AIX 5.3 IY93819	csm.server 1.6.0.2 csm.pessl 1.6.0.0 csm.pe 1.6.0.0 csm.ll 1.6.0.0 csm.ivm.server 1.6.0.0 csm.ivm.client 1.6.0.0 csm.hc_utils 1.6.0.1 csm.hams 1.6.0.2 csm.gui.websm 1.6.0.0 csm.gui.dcem 1.6.0.0 csm.gpfs 1.6.0.0 csm.essl 1.6.0.0 csm.dsh 1.6.0.2 csm.diagnostics 1.6.0.0 csm.deploy 1.6.0.2 csm.core 1.6.0.0 csm.client 1.6.0.0 csm.bluegene 1.6.0.0 RPMS: expect 5.32-1 openCIMOM 0.8-1 openssl 0.9.7d-2

151	5765E88AP	AIX 5.3 IY93818 AIX 5.2 IY93818	AIX-rpm 5.2.0.40-1 tcl 8.3.3-1 tk 8.3.3-1 conserver 8.1.7-2 csm.bluegene 1.5.1.1 csm.client 1.5.1.2 csm.core 1.5.1.3 csm.deploy 1.5.1.2 csm.diagnostics 1.5.1.0 csm.dsh 1.5.1.4 csm.essl 1.5.1.0 csm.gpfs 1.5.1.0 csm.gui.dcem 1.5.1.0 csm.gui.websm 1.5.1.0 csm.hams 1.5.1.1 csm.hpsnm 1.4.1.16 csm.ll 1.5.1.0 csm.pe 1.5.1.0 csm.pessl 1.5.1.0 csm.server 1.5.1.4 RPMS: expect 5.32-1 openCIMOM 0.8-1 openssl 0.9.7d-2 AIX-rpm 5.2.0.40-1 tcl 8.3.3-1 tk 8.3.3-1 conserver 8.1.7-2 csm.client 1.4.1.13 csm.core 1.4.1.15 csm.diagnostics 1.4.1.11 csm.dsh 1.4.1.15 csm.gui.dcem 1.4.1.12 csm.gui.websm 1.4.1.0 csm.hams 1.4.1.11 csm.hpsnm 1.4.1.16 csm.server 1.4.1.15 RPMS: expect 5.32-1 openCIMOM 0.8-1 openssl 0.9.7d-2 AIX-rpm 5.2.0.40-1 tcl 8.3.3-1 tk 8.3.3-1
141	5765E88AP	AIX 5.3 IY93817 AIX 5.2 IY93817	

	246	5765F07AP	AIX 5.3 IY93821	conserver 8.1.7-2 devices.chrp.base.ServiceRM 1.3.0.50 rsct.basic.hacmp 2.4.6.0 rsct.basic.rte 2.4.6.2 rsct.basic.sp 2.4.6.0 rsct.compat.basic.hacmp 2.4.6.0 rsct.compat.basic.rte 2.4.6.0 rsct.compat.basic.sp 2.4.6.0 rsct.compat.clients.hacmp 2.4.6.0 rsct.compat.clients.rte 2.4.6.0 rsct.compat.clients.sp 2.4.6.0 rsct.core.auditrm 2.4.6.0 rsct.core.errm 2.4.6.0 rsct.core.fsrm 2.4.6.0 rsct.core.gui 2.4.6.0 rsct.core.hostrm 2.4.6.1 rsct.core.lprm 2.4.6.0 rsct.core.rmc 2.4.6.2 rsct.core.sec 2.4.6.2 rsct.core.sensorrm 2.4.6.1 rsct.core.sr 2.4.6.0 rsct.core.utils 2.4.6.2 rsct.opt.storagerm 2.4.6.2
RSCT	2310	5765F07AP	AIX 5.2 IY93820	devices.chrp.base.ServiceRM 1.2.0.95 rsct.basic.hacmp 2.3.10.0 rsct.basic.rte 2.3.10.2 rsct.basic.sp 2.3.10.0 rsct.compat.basic.hacmp 2.3.10.0 rsct.compat.basic.rte 2.3.10.0 rsct.compat.basic.sp 2.3.10.0 rsct.compat.clients.hacmp 2.3.10.0 rsct.compat.clients.rte 2.3.10.0 rsct.compat.clients.rte 2.3.10.0 rsct.compat.clients.sp 2.3.10.0 rsct.core.auditrm 2.3.10.0 rsct.core.errm 2.3.10.0 rsct.core.fsrm 2.3.10.0 rsct.core.gui 2.3.10.0 rsct.core.hostrm 2.3.10.1 rsct.core.lprm 2.3.10.0 rsct.core.rmc 2.3.10.2 rsct.core.sec 2.3.10.0 rsct.core.sensorrm 2.3.10.0 rsct.core.sr 2.3.10.0

Parallel ESSL	330	5765F8400	AIX 5.3	rsct.core.utils 2.3.10.2 rsct.opt.saf.amf 2.3.10.1 rsct.opt.storagerm 2.3.10.2 rsct.crypt.rsa512 2.4.5.0 rsct.crypt.rsa1024 2.4.5.0 rsct.crypt.des 2.4.5.0 rsct.crypt.aes256 2.4.5.0 rsct.crypt.aes128 2.4.5.0 rsct.crypt.3des 2.4.5.0 pessl.rte.common 3.3.0.0 pessl.rte.hv 3.3.0.0 pessl.rte.rs1 3.3.0.0 pessl.rte.smp 3.3.0.0 pessl.man.en_US 3.3.0.0 pessl.rte.mp 3.3.0.0 pessl.rte.rs2 3.3.0.0 pessl.rte.up 3.3.0.0 pessl.loc.license 3.3.0.0
	320	5765F8400	AIX 5.3 PK21664 AIX 5.2 PK21664	pessl.rte.common 3.2.0.1 pessl.rte.hv 3.2.0.1 pessl.rte.rs1 3.2.0.1 pessl.rte.smp 3.2.0.1 pessl.man.en_US 3.2.0.1 pessl.rte.mp 3.2.0.0 pessl.rte.rs2 3.2.0.0 pessl.rte.up 3.2.0.0 pessl.loc.license 3.2.0.0
ESSL	420	5765F8200	AIX 5.3 PK19344 AIX 5.2 PK19344	essl.rte.common 4.2.0.4 essl.rte.rs1 4.2.0.4 essl.rte.rs2 4.2.0.0 essl.rte.smp 4.2.0.4 essl.rte.mp 4.2.0.0 essl.rte.up 4.2.0.0 essl.man.en_US 4.2.0.0 essl.loc.license 4.2.0.0

Recommended installation sequence (overview)

The following overview lists the recommended installation sequence for installing software, firmware and operating system updates.

1. HMC software

Important Preliminary steps

- Verify that the HMC is at HMC V3R3.0 or higher before you install this update (required).
- Check if a BIOS update is required on the HMC.
- Check if BIOS hyperthreading is to be disabled on the HMC.
- Disable the SNM/FNM software from the Switch Network Management Panel.

Perform one of the following installation tasks:

- New Install HMC from Recovery CD *or*
-
- Install Upgrade from Recovery CD *or*
-
- Update from .zip file (web)
-

Level Check: Verify that the HMC Code Level is the current level shown in Component update/download information.

2. SNM software
For detailed information, link to the specified SNM ReadMe from the HMC page under the SNM tab.

For a thorough Discussion of the SNM GUI Control panel , please see: Appendix E. Switch Network Manager (SNM) in pSeries High Performance Switch Planning, Installation, and Service

Level Check: Verify that the SNM Code Level is the **current level** shown in Installation guidelines, SNM Component update/download information.

3. HPS/SNI LP software
IMPORTANT:

If you are upgrading from Service Pack 6 or below - do not reboot logical partitions (LPARs) until after Step 4 Install GFW is complete!

See Problem #1 in the "Known Problems" section of this document for more information.

Level Check: Verify that the HPS Code Level is the current level as shown inthe Component update/download information.

4. GFW software
Using the recommended AIX command line (update_flash) method with a locally available GFW img file.

Level Check: Verify that the GFW Code Level is the Component update/download information.

For detailed download and unpacking procedures, see the FAQ for Microcode downloads.

5. Install Power Subsystem Microcode on each frame
Level Check: Verify that the Power Code Level is the current level shown in Component update/download information.

6. Install AIX base updates and any PTFs on each node
Level Check: See AIX download info for the recommended AIX service level for this Service Pack.

7. Install AIX LP updates on each node

Note:

CSM LP's need to download and install the openCIMOM-0.8-1 RPM update.

If you are APPLYING software for csm.server 1.4.1.1, please install the following images from the AIX Installation Media Volume 2:

- tcl
- tk
- expect
- conserver-8.1

Level Check: see "Detailed LP Level Check"

Installation guidelines

Install the software in the order listed in this section.

HMC

Installing HMC software consists of performing some important preliminary steps, and then choosing one of the three available methods for installing the HMC software itself.

Perform one of the following installation tasks:

- New Install HMC from Recovery CD OR
- Install Upgrade from Recovery CD OR
- Update from .zip file (web)

The HMC installed with this Service Pack has an updated HMC WebSM PC CLIENT. To reinstall this version:

Uninstall - reinstall HMC WebSM PC CLIENT (Install Shield version)

Level Check: Verify that the HMC Code Level is the current level shown in Component updates/ download information section.

Important preliminary steps

1. Verify HMC Code level

   Installation of HMC Recovery CD requires an upgrade install if you are installing an existing HMC whose version is R3 V2.6 or less.

   Please read HMC corrective service for important information regarding HMC Recovery CDs.

   Contact your IBM Sales Representative or Business Partner to order HMC Machine Specific CDs
2. Check if BIOS update is required on HMC. There is a mandatory BIOS upgrade required for these HMC PC's:

   7315-C03, 7310-C03, 7315-CR2, 7310-CR2

If your HMC model is not listed, skip this step. If you are updating the HMC on a listed model, then you must first update the BIOS of that HMC model. The BIOS and install instructions can be obtained by linking to the "BIOS Updates" for the referenced machine from the HMC POWER4 servers page. This BIOS will also ship as part of Feature Code 0960.

3. Check if BIOS hyperthreading is to be disabled on HMC.

   Many of the rack mounted HMC's (8187-KUH, 7315-C03) have a BIOS option to enable hyperthreads. The imbedded kernel will not run well when this option is enabled. You must disable this setting before upgrading to HMC3.3.5.
4. Decide whether to install, upgrade or update.

   Beginning with HMC Version 3 Release 3.2, (Service Pack 10) the Ext3 (JFS) filesystem will be enabled if customers perform a New Install/Upgrade to this new level of code by using the HMC Recovery CDs. The Ext3 filesystem is a journaled filesystem and is more reliable and less prone to corruption in case of unexpected loss of power on the HMC.

   Please note the difference between Upgrade and Update:

   - **Upgrade** is done via cd load as described in Perform a New Install/Upgrade.
   - **Update** is done via downloading a zip file as described in Perform an Update.

   Notes:

   Updating to HMC Version 3 Release 3.3 or later using the Install Corrective Service will not enable this feature.

For an Upgrade installation, the following steps must be taken before rebooting the HMC for the Upgrade process:

1. Ensure that the user's home directories are not filled up with debug data. The Upgrade partition only has 2GB in free space to preserve the upgrade data.
2. Ensure that HMC debug is turned off by running the **pedbg -d off** command.

   With debug enabled, certain log files are locked for writing and prevent the Save Upgrade Data task from completing.

3. Perform the Save Upgrade Data task from the Software Maintenance Panel on the HMC console. This task should be run immediately before rebooting the HMC with volume 1 of the recovery CD. If the HMC reboot does not go to the install menu of volume 1 of the recovery CD, you should repeat the save upgrade task.

The procedure for both Installation and Upgrade is identical except:

- For New Installation: When asked to perform an Install/Recovery or Upgrade, select Install/ Recovery F8.
- For Upgrade Installation: When asked to perform an Install/Recovery or Upgrade, select Upgrade F1.

Installation steps

1. Disable the SNM/FNM software from the Switch Network Management Panel.

   Disable SNM Software on **ALL** HMC's attached to the cluster. For each HMC attached to the cluster, select the "Disable SNM Software" option to open a task dialog box that stops the SNM daemon for both verification mode and normal mode.

   See: Appendix E. Switch Network Manager (SNM) in pSeries High Performance Switch Planning, Installation, and Service.

2. Perform a New Install/Upgrade:

   The HMC Recovery CD package is now a set of 2 CDs. Follow these steps to upgrade:

   HMC Recover CD Version 3 Release 3.5

   a. Reboot the HMC with volume 1 of the recovery CD inserted in the DVD Ram drive.

      NOTE: If the HMC fails to boot volume 1 of the recovery CD, the boot sequence in the HMC BIOS may need to be changed so that the DVD/CDROM is before the hard disk in the startup sequence. If you have run the save upgrade data task before the startup sequence was set correctly, then you should rerun the save upgrade data task before installing the HMC with volume 1 of the recovery CD.

   b. Select F8 for New Installation

      OR

      Select F1 for an Upgrade installation. (NOTE: This is NOT the same as an UPDATE)

   c. On the next screen, select F1 to confirm your selection.

      The Install/Upgrade process proceeds until you are prompted to insert the second CD.

   d. Remove the CD from the DVD Ram drive and hit enter when the install is completed.

3. Perform an Update.

   If you are UPDATING from Service Pack 7 (HMC V3 R3.0) or higher and choose to do the UPDATE, then install the following HMC PTF:

   HMC_Update_V3R3.7.zip

   You can download this update from this location: HMC Version 3.3.X corrective service

   Install from the HMC support link ONLY if the HMC is at a Release 3 Version 3.X level (3.3.X):

   1. Select the HMC_Update_V3R3.7.zip link to download the update package.
   2. Use the HMC --> Install Corrective Service option to install.
   3. Reboot HMC after successful installation.

   You may install this UPDATE directly from the web via "HMC Install Corrective Service". See

Step 3, Install Update, in the HMC update for V3R3.7 Readme.

At the HMC interface, follow these steps to install the update:

1. Select Software Maintenance
2. Select Install Corrective Service
3. Select HMC
4. Select Download corrective service from remote system
5. Enter the specified information in the following fields:

   Remote Site: ftp.software.ibm.com
   Patch File: /software/server/hmc/updates/HMC_Update_V3R3.7.zip
   User ID: anonymous
   Password: <your email address>

The HMC interface retrieves the update package from the remote FTP server and begins the install process.

Reboot the HMC after the installation of the update has completed. Rebooting ensures that all changes are available immediately.

After the HMC is rebooted, to Verify a successful update:

1. Select Help in the top menu bar.
2. Select About the Hardware Maintenance Console for pSeries.
3. On the "About" splash panel, check for the following information:

The Version is 3
The Release is 3.7

Web-based System Manager Remote Client (WebSM PC Client)
After the HMC Install New/Upgrade or Update is complete, the Install Shield version of the Web-based System Manager Remote Client (WebSM PC CLIENT) may need to be reinstalled on your remote server or PC. Follow these instructions:

1. Uninstall an existing WebSM client:

   Use the "Add/Remove Programs" option on the Windows Control Panel. Select "Web-Based System Manager Remote Client, and then click the Change/Remove button.
2. Install a new WebSM client:

   Use your preferred web browser to visit the recently installed HMC at this location:

   http://<hmc-hostname>/remote_client.html

   This download and the installation may take ten to twenty minutes.

For pertinent information see the WebSM discussion in the Software updates section of this Readme.

For complete details refer to Chapter 9, Installing and Using the Remote Client, in the Hardware Management Console for pSeries Installation and Operations Guide. (SA38-0590-07)

HMC level check
Check the HMC level as follows:

On the command line:

The output of the **lshmc -V** shows:
Version: 3
Release: 3.7
HMC Build level 20060207.1

On the GUI:

Displaying the spash panel by selecting **Help > About Hardware Management Console** from the menu shows:
Version: 3
Release: 3.7
HMC Build level 20060207.1

HMC Important Notes
Install the HMC code by following the instructions in the HPS Guide. Have your IBM CE download the most recent copy of the HPS guide from IBM CORE to get updated HPS install information. You should also review the HMC information on the web page where you downloaded the images.

This Service Pack REQUIRES:

HMC V3.3.0 (required since Service Pack 6 release)

- HMC V3.3.0 is a NEW BASE release of the HMC introduced in Service Pack 6 that uses a new imbedded kernel.
- This version MAY OVERWRITE root directories (for /, /home/root and /home/hscroot) deleting any scripts that may be there.
- This version MAY delete the Power Subsystem Microcode RPM images on your HMC and you will have to reaquire it for future installs.

The HMC is now installed using two CDs.

The login available at virtual console 0 (via the CTRL-ALT-F1 key sequence) is no longer available.

New Installation and PTF update installation are supported for this release.

Upgrade installation is only supported when upgrading from HMC 3.2.X or greater.

As part of any system change, it is recommended that you maintain a hard copy of network connections, 8 port/ran box configurations and Switch Group IP's.

- Network connections are on the GUI:
  HMC Maintenance Panel => System Configuration => Customize Network Settings:
  IP Address and Netmask for Ethernet0 and Ethernet1, Default Gateway, Nameserver, Domain
- 8 port RAN box configurations are on the GUI:
  HMC Maintenance Panel => System Configuration => Configure Serial Adapter:
  Option 2 shows the current configuration.
- Switch Group IP's are on the GUI: Switch Management => Switch Utilities => Switch Group Configuration

Known problems and issues with the HMC V3 R3.0 Environment:

- wu-ftp will be removed from the HMC distribution. One will be able to ftp out of the HMC but not into the HMC. The 'scp' command is available if you enable secure shell (ssh).
- The websm PC client has a performance decline when downloading the plugin classes from the server. The first time an operation is performed using the client, the task may be slow to launch. Subsequent use of the task, will respond as normal.

Retain Tip on how to use pesh:

To give IBM support personnel the ability to retrieve certain trace/debug information on the HMC, the customer should create a user "hscpe" and assign a password. IBM support can contact the customer to get the password, and then remotely connect to the HMC (with customer consent).

This allows IBM support to perform additional functions, such as viewing logs or starting trace to diagnose problems on the HMC. This user has access similar to the "hscroot" user on HMC. When accessing the HMC remotely via ssh, the "hscpe" user is put into the restricted shell environment. To bypass the restricted shell, the pesh command is provided. The pesh command can only be run by the "hscpe" user, allowing this user to pass in the serial number of the HMC. If the serial number is correct, the user is required to enter a password obtained from IBM Support. If the password is correct, then the user is then put into the un-restricted shell as user "hscpe".

This allows IBM support to perform additional functions, such as viewing logs or starting trace to diagnose problems on the HMC. This user has access similar to the "hscroot" user on HMC. When accessing the HMC remotely via ssh, the "hscpe" user is put into the restricted shell environment. To bypass the restricted shell, the pesh command is provided. The pesh command can only be run by the "hscpe" user, allowing this user to pass in the serial number of the HMC. If the serial number is correct, the user is required to enter a password obtained from IBM Support. If the password is correct, then the user is then put into the un-restricted shell as user "hscpe".

Example:

pesh 23A345K

Enter the serial number in upper case letters.

You will be prompted for a password. Enter password that was provided by IBM Support in lower case

letters.

The HMC serial number can be queried using the command, "lshmc -v | grep SE" or read from the label that is on the front of the HMC. Use the command "date" to verify that the date of the HMC is for the day you intend to use the pesh command.

Starting with HMC Version 3 Release 3.0 and Version 4 Release 1.0, user can also access the restricted shell terminal on the local HMC, by right mouse click on the desktop and selecting the Terminal--rshterm task. If one login at the HMC as user hscpe, the pesh command can also be run from the restricted shell terminal.

For HMC Version 3 Release 3.0 and below, the "hscpe" user id can be created with any role, however, in order to use some of the High Performance Switch (HPS) debug commands, the Service Rep role needs to be selected.

For new HMC installation(s) follow the instructions as described in IBM Hardware Management Console for pSeries Installation and Operations Guide.

To understand how to connect the rs422/rs232 cables see the HPS Guide: Chapter 6; Step 6, "Install the Hardware Management Console (HMC)" thru Step 16. "Verify Installation is Complete"

For "Code load requirements for existing server frames" see Chapter 6

For p655 "Code load requirements for existing p655 server frames"

For p690 "Code load requirements for existing p690 server frames"

SNM/FNM Installation Guidelines
Install SNM Software

Follow these steps to install SNM software:

1. Disable SNM Software on ALL HMCs attached to the cluster. The SNM/FNM software should already be disabled as part of the installation procedure for HMC software. If the SNM software is not disabled, disable it now.
2. Install SNM update, required on all of the HMCs in the cluster. (The version of SNM distributed with the HMC code is not the most current version.)

   To install the corrective service directly from the Internet, use the HMC Corrective Service GUI as follows:

   Remote Site: **ftp.software.ibm.com**
   Patch File: **/software/server/hmc/fixes/<SNM Update zip file>**
   User ID: **anonymous**
   Password: **<your email address>**

   The HMC GUI retrieves the update package and begins the install process.
3. Reboot all HMCs to complete SNM software update.

Note: Note: Do not reboot HMCs until the corrective service has been successfully installed on all HMCs.

After the HMC is rebooted verify that the corrective service update was successful.
DO NOT ENABLE SNM SOFTWARE AT THIS TIME

Level Check:

Select Switch Network Management from the Switch Management folder in the Navigation area.
The Status line, the last line on the Switch Network Management panel, should show:
Current STATUS RPM Version.

Management Properties > Management > 'SNM Version' column should show:
Current SNM Version. 'SNM Version'

NOTES:
Refer to the pSeries High Performance Switch (HPS) Planning, Installation and Service Guide (GA22-7951-01) (HPS Guide) for more details on the Switch Network Manager. Review the following chapters:
Chapter 4. System management components
Chapter 6. Installation: Bringing the network online: Step 1: Enable SNM
Appendix E. Switch Network Manager (SNM) - The SNM Graphical User Interface

The SNM GUI does NOT update its view automatically. You MUST refresh the display via one of the following:

- The GUI menu "Reload" button
- The "Menu">"View">"Reload" function
- The 'F5' key

HPS/SNI LP Installation Guidlines
Install HPS/SNI LP Software.

1. Application of AIX provides base level HPS. Apply HPS/SNI LP Base fileset images to LPARs. (See AIX Licensed Products in the **Component udpates** section.

   Leave SNM Software disabled. To verify the service levels for HPS/SNI LPs, on each LPAR issue:
   lslpp -Lc | egrep "HPS|sni" | cut -d : -f 2,3 |sed 's/:/ /'
   or
   dsh "lslpp -Lc | egrep \"HPS|sni\" | cut -d : -f 2,3 |sed 's/:/ /' "|dshbak|more

   Refer to Section 3 - Detailed LP Level Check for correct levels.

   If you are upgrading from Service Pack 6 or lower **DO NOT** reboot logical partitions (LPARs). See Note 1. Go to Step 2.

   If you are upgrading from Service Pack 7 or higher, verify that the 64 bit kernel is currently in use. Verify the TLP settings. See Note 2.

Reboot LPARs. You should reboot the LPARs as soon as possible to properly integrate the changes and to avoid disruption of current functionality. Go to the Install GFW procedure.

Notes:

1. If you are upgrading from Service Pack 6 or below, do not reboot logical partitions (LPARs) until GFW installation is complete. Rebooting prematurely will generate "phantom" SNI devices. See the **Known problems and workarounds** section for more information.
2. If you are upgrading from Service Pack 7 or higher you should already have set up LPARs to boot the 64 bit kernel and enabled the Technical Large Page (TLP) option, as described in the next step.

   Verify 64 bit kernel is currently in use (on an LPAR): bootinfo -K

   64

   Verify the TLP settings:

   vmo -a|grep lg

   lgpg_size = 16777216
   lgpg_regions=YYY where YYY is the number of technical large pages to export
   (Ex: lgpg_regions = 256)
   soft_min_lgpgs_vmpool = 0
3. It is strongly recommended that users read and become familiar with the items covered in HPS/SNI Notes.
4. IBM pSeries HPS now requires that you set up LPARs with 64 bit kernel and enable Technical Large Page (TLP) option.

   After successful installation of HPS Filesets from levels at Service Pack 6 or lower, 64 bit kernel and technical large page support option must be enabled.

To set up your LPARs with 64 bit kernel:

1. Check which kernel is currently in use: bootinfo -K
   a response of "32" is a 32bit Kernel
2. ln -fs /usr/lib/boot/unix_64 /unix
3. ln -fs /usr/lib/boot/unix_64 /usr/lib/boot/unix
4. Determine which rootvg hdisk contains the boot logical volume (usually hd5). This hdisk will be your "ipldevice".
   a. lspv |grep rootvg
      ```
      hdisk0 009b982332a1f9b8 rootvg active
      hdisk1 009b982332a2321a rootvg active
      ```
   b. lspv -l hdisk0 |grep hd5
      ```
      hd5 1 1 01..00..00..00..00 N/A
      ```
      (hdisk0 is your ipldevice)
5. Issue: bosboot -ad /dev/<ipldevice>
   (eg. bosboot -ad /dev/hdisk0)
6. Reboot: shutdown -Fr
7. Verify 64 bit kernel is running after reboot:

bootinfo -K
64

Regarding the use of memory by TLP

Also refer to Regarding the use of memory by RDMA and TLP in the **Software updates** section.

To set up Large Page Option:

For configuration details, see Large Page Support in the following publication:

AIX 5L Version 5.2 Performance Management Guide

The number of TLP depends on customer configuration and relates to the number of windows required for each adapter(sni) plus any Large Pages page used by other applications.

Set up Large Page Option using the vmo command for each node or node group:

vmo -r -o v_pinshm=1 -o lgpg_size=16777216 -o lgpg_regions=YYY
where YYY is the number of Technical Large Pages to export.

For Example:
To set up a node with 8 sni adapters:
16MB Large Page: lgpg_size = 16777216
256 Large Pages: lgpg_regions = 256
dsh <nodelist> "echo y|vmo -r -o v_pinshm=1 -o lgpg_size=16777216 -o lgpg_regions=256"

Use "echo y|vmo", otherwise vmo will prompt for verification to run bosboot)
Would generate this response:

```
Setting v_pinshm to 1 in nextboot file
Setting lgpg_size to 16777216 in nextboot file
Setting lgpg_regions to 256 in nextboot file
Warning: some changes will take effect only after a bosboot and a reboot
Run bosboot now?
bosboot: Boot image is 19624 512 byte blocks.
Warning: changes will take effect only at next reboot
```

NOTE: The vmtune sample program is being phased out and is not supported in future releases. It is replaced with the vmo command (for all the pure VMM parameters) and the ioo command (for all the I/O related parameters) which can be used to set most of the parameters that were previously set by vmtune. The -v flag has been added to vmstat to replace the -A flag which display counter values instead of tuning parameters. For AIX 5.2, a compatibility script calling vmo and ioo is provided to help the transition.

To Check that Large Page Option is set:
vmo -a|grep lg

```
lgpg_size = 16777216
```

```
lgpg_regions = YYY <where YYY is the number of Technical Large Pages to export>
soft_min_lgpgs_vmpool = 0
```

HPS/SNI Notes
Notes on using TLP (Large Page) Settings in an HPC environment

It is strongly recommended that users familiarize themselves with TLP basics and configuration options available to them, at this location. Federation switch adapter requires TLP usage and these TLP requirements are documented (see "Here is a formula to calculate the required TLP" in the HPS/SNI LPs section below) in a latter section of this document.

You should also consult the section on Large page feature on AIX in the following publication:

The AIX 5L Version 5.2 Performance Management Guide

NOTE: Users need to be aware of the usage of the LoadLeveler pre-emption features with TLP (Large Pages), specifically that jobs that are using TLP that are pre-empted will essentially "lock up" the real memory the TLP's use, which is pinned by AIX. Unwise use of TLPs with LoadLeveler pre-emption can result in exhausting real memory available for jobs.

If you want LoadLeveler to schedule jobs based on the availability of large page, (especially if the job is going to run in mandatory Large Page mode) you may consider using the LoadLeveler consumable resource feature. The function has been available for several years and is documented in the LoadLeveler manual.

Notes on tuning Virtual Memory Settings in an HPC environment

Customers should be advised that the AIX VMM parameters (set by the vmo command) minfree and maxfree will most likely have to be adjusted (increased) in an HPC environment based on cluster size, the amount of system memory and the number of processors per CEC. When tuned properly, these settings will ensure enough memory remains available for core cluster infrastructure applications (RSCT, GPFS, LL). The recommended initial value for these tunables are

```
minfree = 10000
maxfree = 12000
```

Users are strongly urged to consult the following AIX documentation on virtual memory and vmstat tools and tune their system accordingly.

http://publib16.boulder.ibm.com/doc_link/en_US/a_doc_lib/aixbman/prftungd/memperf.htm

http://publib16.boulder.ibm.com/doc_link/en_US/a_doc_lib/aixbman/prftungd/

memperf1.htm#i50853

The AIX 5L Version 5.2 Performance Management Guide should also be consulted.

NOTE: Tuning these settings helps you avoid conditions where core cluster applications shut down and restart due to extensive blockage caused by "out of memory" issues. Keep in mind that all cluster applications should be designed and cluster tuned accordingly as to avoid oversubscribing to the real memory available.

GFW Installation Guidelines
The recommended installation is the AIX command line method - update_flash - using locally available GFW img file.

Important Preliminary Notes

Notes on updating GFW code (system firmware) from the AIX command line
Document Reference: pSeries High Performance Switch Planning, Installation, and Service.

For "Code load requirements for existing p690 and p655 server frames" please see the section in Chapter 6 titled "Step 3: p690 GFW code load" or "Step 3: p655 GFW code load" respectively in the HPS guide.

To understand how to connect the rs422/rs232 cables see the HPS Guide:
Chapter 6; Step 6, "Install the Hardware Management Console (HMC)" thru Step 16. "Verify Installation is Complete"

For "Code load requirements for existing server frames" see Chapter 6
For p655 "Code load requirements for existing p655 server frames"
For p690 "Code load requirements for existing p690 server frames"

For each CEC on which you want to install the GFW code, one partition running AIX must have "Service Authority" set. Linux does not support microcode download at this time.

The "Service Authority" is set on one LPAR per CEC in the LPAR's profile "other" tab. This designates the LPAR as authorized to provide update images to the CSP. All partitions except the one with "Service Authority" must be shut down.

The partition with "Service Authority" must own the device from which the microcode update image will be read. It is also recommended that the partition with "Service Authority" have a hard disk.

If the required devices are not in the partition with "Service Authority", the customer or system administrator must reassign the appropriate resources to it. This requires rebooting the partition with "Service Authority".

If the firmware on a full system partition is being updated, no special steps are required to perform the firmware update using the service aid.

Ensure the GFW image file is not corrupted/truncated before you begin the update_flash process.

Check that /var and /tmp directories are not above 50% full on the partition with the service authority.

The update process can range from 20 minutes to 2 hours, depending on system configuration. The system reboots itself during the update process. Since SNM is disabled during this process, the SNI adapter interfaces will NOT be configured, or will be incorrectly configured when the LPAR(s) reactivate.

It is recommended that you use the `update_flash -f` command as opposed to the shutdown -Fu method.

AIX52 APAR IY49146 is required for update_flash to work correctly. Level Check by running this command on the partitions:

```
instfix -ik IY49146
```

The `update_flash` command will reboot the CEC(s) and will activate the LPAR(s).

You may find some more detailed instructions provided on the website with the latest image: pSeries 690 and 670 Version 3 Firmware Update

To install GFW update using Diskette method:
For p690 systems follow the instructions in the HPS guide on "Step 3: p690 GFW (system firmware code load)" in chapter 6.

To install GFW update using NIM method:
For a p655 CEC via NIM, follow the HPS guide Chapter 6, "Code load requirements for existing p655 server frames", Step 3. GFW (system firmware) code load.

Installation steps

This procedure employs the recommended AIX command line (update_flash) method)

1. SNM Software should be still disabled. Verify from the SNM GUI Panel.
2. For each CEC on which you want to install GFW code, shut down all partitions except the one with service authority.
3. Install the appropriate GFW driver on each CEC to be upgraded:

   GFW
   On the AIX partition with Service Authority:
   Copy the GFW firmware update code to /tmp
   Enter the following command:
   /usr/lpp/diagnostics/bin/update_flash -qf /tmp/<gfw img file>
   The system will apply the new firmware, reboot, and return to the AIX prompt.

   If you use dsh to invoke update_flash then use the -q flag so it does not put out a prompt.
   Ex. dsh /usr/lpp/diagnostics/bin/update_flash -qf /tmp/<GFW img file>
   More conveniently, the '?' represents a single character 'wild card' and will select either /tmp/3H050405.img or /tmp/3J050405.img. Do not have both on the same LPAR.)

dsh [-N <nodegrp>] "echo \"/usr/lpp/diagnostics/bin/update_flash -qf tmp/<GFW img file>\"|at now"

```
job root.1102366157.a at Mon Dec  6 15:49:17 2004
job root.1102366120.a at Mon Dec  6 15:48:40 2004
job root.1102366242.a at Mon Dec  6 15:50:42 2004
```

4. After the LPAR(s) are Running, power OFF the CEC(s) from the GUI or by using CSM rpower - not from EPO red switch.
5. After the CEC(s) are powered off, follow these steps to enable the SNM/FNM Software from the GUI:
   1. Select Switch Network Management from the Switch Management folder in the Navigation area.
   2. Select Logical Topology on a 'per plane' basis.
      Select the number of Planes and select the Logical Topology on a per plane basis (Number of Endpoints on a plane) .
      Ex: 2 Frames and 2 Switches with NO Switch-to-Switch Links is 2 planes;

1NSB_0ISB_16EP (16 Endpoints)
Ex: 2 Frames and 2 Switches with ANY Switch-to-Switch Links is 1 plane;
2NSB_0ISB_32EP (32 Endpoints)

3. Enable SNM Software for Normal Operation

   OR

   Enable SNM Software for Switch Network Verification.
4. After Enable SNM Software task completes, verify current HMC and SNM version as needed via the Management Properties task:

   HMC version is under the "Version" tab.
   SNM version is under the Management tab, "SNM Version" column.
   This should show the current version as given in Component updates/download information section.

Notes:

The "Enable SNM Software for Switch Network Verification" task is used during new system setup/ installation or after reconfiguration to initialize and check out the system by NOT removing bad links or routes. The main purpose is to discover cable mis-wires> and defective hardware. No Routing is modified or loaded. It is NOT intended to be used in a working environment.

Refer to the HPS Guide for more details on defining the Switch Network Topology: Review - Chapter 4. System management components; Step 1: Enable SNM
Review - Appendix E. Switch Network Manager (SNM) - The SNM Graphical User Interface

The SNM GUI does NOT update its view automatically. You MUST refresh the display via one of the following:

- The GUI menu "Reload" button
- The "Menu">"View">"Reload" function
- The 'F5' key
- Power up CEC(s) from the HMC GUI and activate logical partition(s).
- Enable technical large page support - Required for levels greater than Service Pack 6.
  Note: Refer to "HPS/SNI" section for details on technical large page setup.
- Determining the level of firmware on the processor subsystem
  Firmware level is indicated as: 3xyymmdd.img; where
  x = a firmware designation such as J or H - J = p655, H=p690
  yy = year, mm = month, and dd = day of the release
  similar to 3J060626

  Check the GFW microcode level from either A VTERM to the main SP Menu. This should show the correct level on the top line.

  OR

  You can also check the GFW level from the AIX command line on the active LPAR(s):
  lscfg -vp | grep alter | grep "\.3"
  You should see:

ROM Level (alterable)..... < GFW img level>
- or -
ROM Level (alterable)..... < GFW img level>
where < GFW img level> is the level given in GFW download information.

- Determining the level of HPS adapter microcode
  The HPS adapter microcode (ucode) is shipped as part of the GFW update image.

  Level Check the ucode to make sure you do not have to reinstall GFW.
  To Level Check the ucode:
  From an AIX52 partition, issue:
  /usr/sni/aix52/debugtools/sni_get_ucode_version -l sniz
  From an AIX53 partition, issue:
  /usr/sni/aix53/debugtools/sni_get_ucode_version -l sniz
  where z = sni interface number on your system anywhere from 0 thru 7 (Eg. sni0)
  which can be seen in "netstat -in" output.

  This should show the current version as given in GFW download information.

  Important You will need to reinstall the GFW update if you:

  - Neglected to disable SNM during the GFW update, or
  - Added/replaced an HPS adapter.

  Otherwise, the HPS ucode may not have been applied correctly.

Power Subsystem Microcode Installation Guidelines
Install Power Subsystem Microcode on each frame by following these steps:

1. Download the Power Subsystem Microcode from the Microcode downloads
   site to an FTP server.

   Install via the HMC GUI through the Software Maintenance -> Frame panels.
   Receive Corrective Service
   Install Corrective Service

   If you are upgrading from Service Pack 9 or above, the installation of Power Subsystem
   Microcode rpm is complete. Proceed to Step 4 to level check installed version.
2. If you are upgrading from Service Pack 8 or lower, the new Power Subsystem Microcode requires
   the switches to be recycled for the changes to become effective.
   Power cycle switch(s):
   From the HMC GUI select "Switch Network Management > Switch Topology View"
   For each switch plane
   - select "Selected > Power <Off"
   - refresh GUI to verify power status
   - select "Selected <Power <On"
   - refresh GUI to verify power status
   Repeat this procedure on all switch planes.
3. Recycle SNM daemon using the HMC GUI:

select Switch Network Management > Disable SNM Software
refresh display: Menu > File > Refresh >Enable SNM Software.

Notes:
Recommended procedure for a complete power-cycle of the cluster ("EPOW")
(Cold Boot Procedure)

In the event that power-cycling the HPS switch boards fails, as a last resort the following procedure is suggested:

For a complete discussion of the subject :
See pSeries High Performance Switch Planning, Installation, and Service, Chapter 9, Service procedures, Managed system power on and power off (LPAR reboot)

The 10 minute wait is for switch frames to stabilize.

Suggested Procedure:

- shutdown all the LPARs
- rpower -a cec_off
- stop fnmd
- EPOW off all the CECs and the ISB frame
- start fnmd
- EPOW on the ISB frame and the NSB frames
- wait 10 minutes, EPOW everyone else on
- check for flashing lights on the (optical) risers in the ISBs
- rpower -a cec_on
- tail /var/hsc/log/*Init.log until that stops
- run hps_check.pl
- activate all the LPARs

4. Level Check

After completing the ptcode installation, go to the HMC GUI and verify successful installation:
- Select: Software Maintenance --> Frame --> Install Corrective Service
- Verify that the "Installed Version" matches the version you just installed.

Important: This window may not automatically refresh when installation completes. Manually refresh the window as necessary.

5. Verify HPS is Functioning.

At this point the basic HPS installation is complete. You should now be able to ping over the switch.

A general check is a good ping all script. The HPS Documentation covers this topic in the host based verification tool.
Post Installation Task 2. (pSeries High Performance Switch Planning, Installation, and Service; "Run the host-based verification tools", p116)

If SNM Software is not running then :

The "Switch Management" | Switch Network Management | views : Switch Topology

View and End Point View will not be populated and there is a dialog message indicating that and you will not be able to ping over the switch.

The hps_check.pl file is not available in a closed box without the root password.
It should show the links as Timed and MPA Available :

```
Lpar Name Lpar# Sni# => Adapter# Csp# Cronus# => Frame Cage
Chip Port : Timed? MPA TOD
c661f1rp02 1 0 0 2 5 1 1 5 2 YES YES SLV
c661f1rp02 1 1 1 3 4 4 4 5 2 YES YES BAK>
```

AIX Installation Guidelines
Install AIX base updates on each node by following these steps.

See AIX download information for the recommended AIX service level for this Service Pack.

For complete instructions on the installation procedures please see:
Document Reference: AIX 5L Version 5.2 Installation Guide and Reference

1. Download and install the recommended AIX 5L version service level for this Service Pack.
    1. Go to Quick links for AIX fixes, and then click the "Specific fixes" link for your Release of AIX.
    2. In the "Search By" drop down box Select "APAR or abstract".
    3. In the "Search String" text box enter the specific fix as given in AIX download information, and then click "Go". The search results page displays a text box with the specific fix requested.
    4. Select that package, and then click "Add to my download list".
    5. You can view by clicking "View my download list".
    6. Repeat these steps for any additional fixes or PTFs.
    7. After all the indicated fixes are in the download list click "Continue".
    8. Specify your "Current level" (use oslevel -r command to determine your AIX level).
    9. After "Select a download server", click "Continue", and then follow the instructions for downloading the selected file sets.

Please read the "Memo to Users" on the Download fixes page for installation and fix information. Level check AIX by running this command on the logical partition(s):

lslpp -ha bos.mp64

Verify that filesets are at or above the given in AIX download information for the recommended AIX service level for this Service Pack.

AIX LP Installation Guidelines
Install AIX LP updates on each node by following these steps.

Level check: "Detailed LP Level Check"
Notes: CSM LPs need to download and install openCIMOM-0.8-1 RPM update.

Document Reference: AIX 5L Version 5.2 Installation Guide and Reference
(SC23-4389-03)
IBM RSCT: Administration Guide

As noted in the Installation Guide, in order to use CSM LPs, you need to download and install openCIMOM-0.8-1 RPM update. You can download openCIMOM from the following location:

http://www-1.ibm.com/servers/aix/products/aixos/linux/download.html

Select "Package" OpenCIMOM "Version" 0.8 (5.2) (For AIX 5.2)
For complete details see: Chapter 4. Installing the management server; Step 6. Download Open Source Software of CSM Guide.

Document Reference:
IBM Cluster Systems Management for AIX 5L Planning and Installation Guide
AIX 5L Version 5.2 Installation Guide and Reference
IBM Reliable Scalable Cluster Technology Administration Guide

DPCL is no longer a part of the IBM PE for AIX licensed program. Instead, DPCL is now available as an open source offering that supports PE.

For more information and to download the DPCL open source project go to:

http://oss.software.ibm.com/developerworks/opensource/dpcl

Document Reference:
IBM Parallel Environment for AIX 5L Installation Version 4, Release 1.1
Chapter 1. Introducing PE 4.1.1

1. Download and install the applicable IBM Virtual Shared Disk(VSD) , LAPI, HPS, PPE, LoadLeveler, GPFS, Parallel ESSL, ESSL, CSM and RSCT PTF updates.

   Apply the listed APARs which are needed for this Service Pack.
   They are available from this web site:
   Quick links to AIX fixes

   If you are upgrading to this Service Pack from a service pack that is earlier than SP9, then installing CSM1.4 is required.

   The Maintenance package contains CSM 1.4 which requires RPM update openCIMOM 0.8(5.2).
2. To verify that the service levels for your LP's are current for this Service Pack, on each logical partition issue:
   lslpp -Lc | egrep "vsd|LAPI|HPS|sni|ppe|LoadL|mmfs|rsct|csm|essl|pessl" | cut -d : -f 2,3 |sed 's/:/ /'

Check the levels against those listed in Detailed LP Level Check

3. RSCT Migration Issues:

   See Known problems and workarounds, "hagsglsm is not reporting the local switch membership group"

   For more information, refer to:
   RSCT Administration Guide; Chapter 3. Creating and Administering an RSCT Peer Domain; Migration

The following known problems, workarounds and restrictions exist for this Service Pack:

Known problems and workarounds

Phantom SNI devices may appear after upgrade from pre-Service Pack 7
Users Affected:

Users upgrading systems from SP6 or earlier with existing SNI devices

Problem Description:

Changes in the HPS switch microcode and driver demand that it is absolutely necessary to install the HPS fileset updates and the GFW firmware updates without rebooting logical partitions before the system firmware is successfully installed. Rebooting the LPAR(s) prematurely will cause "phantom" SNI devices to be created on the partition(s).

The failure signature is:

1. There are twice the number of snX and sniX devices as expected in the "lsdev -C|grep sn" output

   where

   X = sni or sn interface number on your system anywhere from 0 thru 7 -- e.g. sni0 or sn0
2. All the sn interfaces are in the Defined state.
3. Upper half of the sniX devices are in "Available" state (higher numbered devices) and lower half of the sniX devices are in "Defined" state (lower numbered devices).
4. All the sniX devices could also be in the "Defined" state.

Common causes of "phantom" sniX devices:

- If you re-boot LPARs after upgrading HPS filesets, but prior to a successful GFW upgrade.
- Failure during GFW upgrade after updating HPS filesets.

These events/scenarios will cause all LPARs on a CEC to reboot without upgrading the GFW via AIX command line method (i.e. update_flash command)

- If you have a corrupted/truncated GFW image file.
- If either /var or /tmp is too full.
- If an LPAR other than the Set Service Authority LPAR is in "Running" state during update_flash.
- If an LPAR other than the Set Service Authority LPAR is used to run the update_flash command.

Note: There may be other factors that cause LPARs to reboot after the HPS fileset is upgraded and before the GFW is successfully updated.

Recovery Procedure:

1. Complete the GFW update, and then verify that the firmware updates on all CEC(s) was successful.
2. Recover the sniX and snX devices. (Note: To recover requires at least 1 reboot. Two reboots are required if the sniX devices are busy.)
    a. Remove all the SNI devices after naking note of any customization to the SNI devices such as num_windows, driver_debug, etc . Use the "lsattr -El sniX" command so that any customization can be re-applied after the recovery procedure.

    NOTE: This procedure resets all values back to the defaults:

    To remove each logical sniX, execute the following:

    for X in 0 1 2 3; do rmdev -d -l sni$X; done

    If the rmdev fails for any devices (e.g. device is busy), then unconfigure the device driver as follows. Otherwise, go to step **b**.

        1. Rename the configuration method for the device as follows:

        mv /usr/sni/aix52/cfgsni /usr/sni/aix52/cfgsni.orig
        2. Reboot each LPAR that failed.
        3. Run the rmdev loop again.
        4. Restore the original configuration methods name as follows:

        mv /usr/sni/aix52/cfgsni.orig /usr/sni/aix52/cfgsni

    b. Remove ONLY the phantom top half of the snX devices. The lower half are real snX devices and have the ipaddr and netmask attributes in the odm. Do not delete these real snX devices.

    For the top HALF of logical snX devices, execute the following:

for X in 2 3; do rmdev -d -l sn$X; done

3. Reboot the LPAR(s) --> shutdown -Fr.
4. Restore any customization to the SNI devices (e.g. num_windows, driver_debug, etc.).

hagsglsm is not reporting the local switch membership group
Component:

rsct

Systems Affected:

All rsct Users at Service Pack 9

Description:

In order to complete the migration of a peer domain and update the active RSCT version to a new level, you must enter the runact command as follows:

runact -c IBM.PeerDomain CompleteMigration Options=0"

This command should be run after every RSCT release upgrade.

For a more complete discussion see: "Avoiding Domain Partitioning When Migrating From RSCT 2.2.1.x or 2.3.0.x" in *IBM Reliable Scalable Cluster Technology Administration Guide*, Chapter 3. "Creating and Administering an RSCT Peer Domain". See the section on Migration.

Restrictions

No switch should be powered off while the SNM software is running

Component:

SNM - Switch Network Management

Systems Affected:

High Performance Switch (HPS) users applying from pre-Service Pack 7

Implications:

- A CEC frame with a switch in it CANNOT be EPOWed.
- If a CEC has to be power cycled, power down the CEC and not the frame.
- If a CEC frame with a switch needs to be EPOWed, power down the frame, kill the SNM daemon after 5 minutes, power up the frame and restart the SNM daemon.
- If one or more switches need to be recycled, power down the switches, power them back up and then recycle the SNM daemon after 5 minutes.

Rules for swapping cables for fault isolation
Component:

SNM - Switch Network Management

Systems affected:

All HPS Users applying from pre-Service Pack 7

Description:

Follow these rules when swapping cables for fault isolation:

- Only swap SNI attached cables at the switch ports to which they are attached.
- Do not swap switch to switch cables with other switch to switch cables nor with SNI attached cables.
- Do not swap cables on the SNI ports.
- Do not swap cables between switches.
- If adapters are accidentally "miswired" during the process of swapping cables, recable the adapters to their original positions.

Improved performance is more sensitive to bad links

Component:

HPS/LAPI

Problem Description:

The protocol (MPI or LAPI) will timeout if the job runs on bad links and the link routes are not fixed. If the link failure turns into adapter failure then the job gets terminated.

Solution:

To resolve this issue, monitor Service Focal Point for bad links and fix them.

HPS Cluster recommended LPAR reboot procedure

Component:

HPS/SNI

Systems Affected:

All HPS Users applying from pre-Service Pack 7

Description:

To ensure the HPS switch links are properly shutdown and re-enabled, IBM recommends that you use the following commands to recycle and reboot all LPAR/AIX images in your cluster.

1. To shut down LPAR:

   shutdown -F
2. To reboot LPAR or multiple LPARs simultaneously:

   shutdown -Fr

   OR

   dsh -av shutdown -Fr

If this procedure is not followed, then en masse reboot will certainly result in one IPC1:37:CC MP Fatal event per link.

Use of the "reboot" command or "rpower" commands will not shutdown the HPS switch links in an

orderly fashion - when more than one frame at a time is cycled concurrently the SNM daemon may hang and Service Focal Point could end up with artificial errors. If when these commands must be run concurrently on multiple LPARS, it is recommended that you use them one frame at a time in your cluster.

When the HMC GUI is used to cycle an LPAR, it is recommended that you use the "shutdown" option to recycle the LPAR, not the "reset" option. The "shutdown" option will ensure that the HPS switch links are shutdown and re-enabled cleanly - whereas the "reset" option(s) will essentially use the rpower command (and not do an orderly shutdown).

Note: Use of the rpower or HMC GUI reset options should be reserved as a "last resort" for LPARS that are not responding to a shutdown command.

Striping mode Restrictions

Component:

HPS/SNI

Description:

- Striping is currently limited to **eight** links per task. Multiple tasks can use all links as long as there are more tasks than half the links available.
- Performance degradation occurs for single task per link of 3% for unidirectional and 5% for exchange bandwidth at large messages. However, multiple tasks per link get full link bandwidth.
- In non-striping mode you only get half the links on a p655 system using a 2 plane configuration when setting MP_EUIDEVICE=sn_single. You must use MP_EUDEVICE=csss or sn_all to get all links.