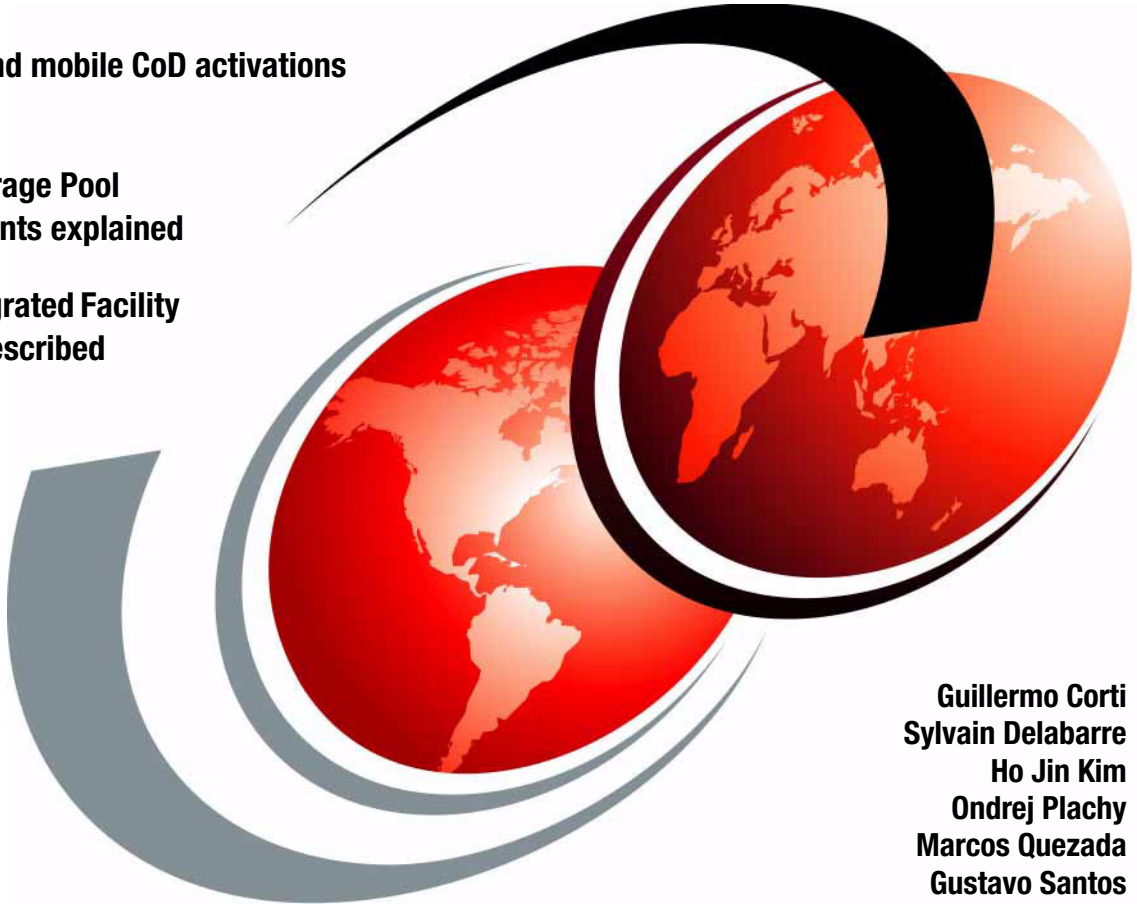
IBM

IBM PowerVM Enhancements
What is New in 2013

- PowerVP and mobile CoD activations explained
- Shared Storage Pool enhancements explained
- Power Integrated Facility for Linux described



Guillermo Corti
Sylvain Delabarre
Ho Jin Kim
Ondrej Plachy
Marcos Quezada
Gustavo Santos

Redbooks

ibm.com/redbooks

International Technical Support Organization

IBM PowerVM 2013 Enhancements

May 2014

SG24-8198-00

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

First Edition (May 2014)

This edition applies to Version 2.2, of IBM PowerVM Express Edition (5765-PVX), IBM PowerVM Standard Edition (5765-PVS), IBM PowerVM Enterprise Edition (5765-PVE), IBM PowerVM EE Edition for Small Server (5765-PVD), IBM PowerVM for Linux Edition (5765-PVL), Hardware Management Console (HMC) V7 Release 7.8.0, IBM Power 740 Firmware Level AL740-110, and IBM Power 750 Firmware Level AL730-122.

Contents

Figures

Tables

Examples

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 grant 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 websites are provided for convenience only and do not in any manner serve as an endorsement of those websites. The materials at those websites are not part of the materials for this IBM product and use of those websites 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.

Any performance data contained herein was determined in a controlled environment. Therefore, the results obtained in other operating environments may vary significantly. Some measurements may have been made on development-level systems and there is no guarantee that these measurements will be the same on generally available systems. Furthermore, some measurements may have been estimated through extrapolation. Actual results may vary. Users of this document should verify the applicable data for their specific environment.

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

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

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

Trademarks

IBM, the IBM logo, and ibm.com are trademarks or registered trademarks of International Business Machines Corporation in the United States, other countries, or both. These and other IBM trademarked terms are marked on their first occurrence in this information with the appropriate symbol (® or ™), indicating US registered or common law trademarks owned by IBM at the time this information was published. Such trademarks may also be registered or common law trademarks in other countries. A current list of IBM trademarks is available on the Web at http://www.ibm.com/legal/copytrade.shtml

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

AIX®
Enterprise Storage Server®
GDPS®
Geographically Dispersed Parallel Sysplex™
GPFS™
HACMP™
IBM®
Parallel Sysplex®
POWER®
POWER Hypervisor™
Power Systems™
POWER6®
POWER7®
POWER7 Systems™
POWER7+™
PowerHA®
PowerLinux™
PowerVM®
PowerVP™
Redbooks®
Redbooks (logo)  ®
Tivoli®

The following terms are trademarks of other companies:

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

Microsoft, Windows, 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 Oracle and/or its affiliates.

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

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

Preface

IBM® Power Systems™ servers coupled with IBM PowerVM® technology are designed to help clients build a dynamic infrastructure, helping to reduce costs, manage risk, and improve service levels.

IBM PowerVM delivers industrial-strength virtualization for IBM AIX®, IBM i, and Linux environments on IBM POWER® processor-based systems. IBM PowerVM V2.2.3 is enhanced to continue its leadership in cloud computing environments. Throughout the chapters of this IBM Redbooks® publication, you will learn about the following topics:

- New management and performance tuning software products for PowerVM solutions. Virtual I/O Server (VIOS) Performance Advisor has been enhanced to provide support for N_Port Identifier Virtualization (NPIV) and Fibre Channel, Virtual Networking and Shared Ethernet Adapter, and Shared Storage Pool configurations. IBM Power Virtualization Performance (PowerVP™) is introduced as a new visual performance monitoring tool for Power Systems servers.
- The scalability, reliability, and performance enhancements introduced with the latest versions of the VIOS, IBM PowerVM Live Partition Mobility, and the Hardware Management Console (HMC). As an example, this book goes through the Shared Storage Pool improvements that include mirroring of the storage pool, dynamic contraction of the storage pool, dynamic disk growth within the storage pool, and scaling improvements.

This book is intended for experienced IBM PowerVM users who want to enable 2013 IBM PowerVM virtualization enhancements for Power Systems. It is intended to be used as a companion to the following publications:

- *IBM PowerVM Virtualization Introduction and Configuration*, SG24-7940
- *IBM PowerVM Virtualization Managing and Monitoring*, SG24-7590

Authors

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

Guillermo Corti is an IT Specialist at IBM Argentina. He has been with IBM since 2004 and has a 20-year technical background on Power Systems and AIX. He has a degree in Systems from Moron University. He also has 10 years experience working in service delivery on Power Systems, AIX, VIOS, Linux, and HMC for North American accounts.

Sylvain Delabarre is a Certified IT Specialist at the IBM Client and Power Systems Linux Center in Montpellier, France. He has been with IBM France since 1988. He has worked as a Power Systems Benchmark Specialist since 2010. He also has 20 years of AIX System Administration and Power Systems experience working in service delivery, AIX, VIOS, and HMC support for EMEA.

Ho Jin Kim is a Senior Power Systems System Service Representative in IBM Korea. He has been with IBM since 2001. He has a Masters degree in Management Information Sciences from Hancock University of Foreign Studies. He has supported Finance Sector accounts for seven years. He is a Red Hat Certified Engineer and his areas of expertise include Power Systems solutions, Linux on Power Systems, AIX, and PowerVM.

Ondrej Plachy is an IT System Engineer in Comparex CZ, s.r.o. He has 18 years of experience in UNIX and storage systems. He holds the Ing. academic degree in Computer Science from Czech Technical University (CVUT), Prague. He has worked at IBM for nine years as part of the IBM AIX software services team. Now, he is responsible for high availability (HA), disaster recovery (DR) project design, implementation, and the support of large-scale data center computer systems in the Comparex company. Comparex is an IBM Business Partner in the Czech Republic.

Marcos Quezada is a Consulting IT Specialist at IBM Latin America Advanced Technical Support for Power Systems. He is an IBM Level II Certified IT Specialist with 15 years of IT experience in several related positions as a Brand Development Manager and as an IT Specialist at IBM Global Techline. He holds a degree in Systems Engineering from Fundación Universidad de Belgrano, Argentina. As a Consulting IT Specialist, he leads IBM teams in promoting IT and POWER Architecture technologies to achieve buy-in from clients. He consults on virtualization, cloud computing, and big data.

Gustavo Santos is an IT Architect at IBM Delivery Center Argentina. He has been with IBM since 1997. He has 17 years of experience in Power Systems and the UNIX field. He holds a degree in Systems Engineering from Universidad Abierta Interamericana. He also has 15 years experience working in service delivery on AIX, VIOS, and HMC for multiple accounts in the US and LA. His areas of expertise include Power Systems, AIX, VIOS, and UNIX.

The project that produced this publication was managed by:
Scott Vetter, PMP

Thanks to the following people for their contributions to this project:

Syed R Ahmed, Suman Batchu, Carl Bender, David Bennin, Bill Casey, Ping Chen, Shaival Chokshi, Rich Conway, Michael Cyr, Robert K Gardner, Yiwei Li , Nicolas Guérin, Stephanie Jensen, Manoj Kumar, P Scott McCord, Nidugala Muralikrishna, Paul Olsen, Steven E Royer, Josiah Sathiadass, Vasu Vallabhaneni, Bradley Vette.

Now you can become a published author, too!

Here's an opportunity to spotlight your skills, grow your career, and become a published author—all at the same time! Join an ITSO residency project and help write a book in your area of expertise, while honing your experience using leading-edge technologies. Your efforts will help to increase product acceptance and customer satisfaction, as you expand your network of technical contacts and relationships. Residencies run from two to six weeks in length, and you can participate either in person or as a remote resident working from your home base.

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

ibm.com/redbooks/residencies.html

Comments welcome

Your comments are important to us!

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

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

  ibm.com/redbooks

- Send your comments in an email to:

  redbooks@us.ibm.com

- Mail your comments to:

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

Stay connected to IBM Redbooks

- Find us on Facebook:

  http://www.facebook.com/IBMRedbooks

- Follow us on Twitter:

  http://twitter.com/ibmredbooks

- Look for us on LinkedIn:

  http://www.linkedin.com/groups?home=&gid=2130806

- Explore new Redbooks publications, residencies, and workshops with the IBM Redbooks weekly newsletter:

  https://www.redbooks.ibm.com/Redbooks.nsf/subscribe?OpenForm

- Stay current on recent Redbooks publications with RSS Feeds:

  http://www.redbooks.ibm.com/rss.html

1

Introduction

This publication provides a summary of all the major IBM PowerVM and Hardware Management Console (HMC) enterprise enhancements introduced in the October 2013 announcement.

Before you continue, you need to be familiar with and have practical experience with the contents in the following IBM Redbooks publications: *IBM PowerVM Virtualization Introduction and Configuration*, SG24-7940, and *IBM PowerVM Virtualization Managing and Monitoring*, SG24-7590.

This book was written so that you can go through the pages starting here or jump to whatever subject interests you. The following chapters and sections of this book are briefly introduced in this chapter:

- IBM Power Virtualization Center (IBM PowerVC)
- IBM Power Virtualization Performance (IBM PowerVP) for Power Systems
- Power Integrated Facility for Linux (IFL)
- Virtual I/O Server (VIOS) 2.2.3
- VIOS Performance Advisor
- PowerVM Live Partition Mobility
- Hardware Management Console (HMC) feature updates

1.1 IBM PowerVC

IBM Power Virtualization Center (IBM PowerVC) is designed to simplify the management of virtual resources in your Power Systems environment.

After the product code is loaded, IBM PowerVC's no-menus interface will guide you through three simple configuration steps to register physical hosts, storage providers, and network resources. Then, it starts capturing and intelligently deploying your virtual machines (VMs), among other tasks shown in the following list:

- Create VMs and then resize and attach volumes to them.
- Import existing VMs and volumes so they can be managed by IBM PowerVC.
- Monitor the utilization of the resources that are in your environment.
- Migrate VMs while they are running (*hot migration*).
- Deploy images quickly to create new VMs that meet the demands of your ever-changing business needs.

IBM PowerVC is built on OpenStack. OpenStack is open source software that controls large pools of server, storage, and networking resources throughout a data center.

IBM PowerVC is available in two editions:

- IBM Power Virtualization Center Express Edition
- IBM Power Virtualization Center Standard Edition

Table 1-1 shows an overview of the key features included with IBM PowerVC Editions.

Table 1-1 Key features of IBM PowerVC editions

IBM Power Virtualization Center Express Edition	**IBM Power Virtualization Center Standard Edition**
► Supports IBM Power Systems hosts that are managed by the Integrated Virtualization Manager (IVM). ► Supports storage area networks, local storage, and a combination in the same environment. ► Supports a single VIOS VM on each host.	► Supports IBM Power Systems hosts that are managed by a Hardware Management Console (HMC). ► Supports storage area networks. ► Supports multiple VIOS VMs on each host.

For more information about IBM PowerVC, see the *IBM PowerVC Introduction and Configuration*, SG24-8199.

1.2 IBM PowerVP

IBM Power Virtualization Performance (IBM PowerVP) for Power Systems is a new product that offers a performance view into an IBM PowerVM virtualized environment running on the latest firmware of IBM Power Systems. It is capable of showing which virtual workloads are using specific physical resources on an IBM Power Systems server.

IBM PowerVP helps reduce the time and complexity to find and display performance bottlenecks through a simple dashboard that shows the performance health of the system. It can help simplify both prevention and troubleshooting and therefore reduce the cost of performance management.

1.3 Power Integrated Facility for Linux

With Power Integrated Facility (Power IFL), IBM is introducing an enterprise Power Systems offering to consolidate or integrate Linux with AIX and IBM i applications and data in large enterprise servers.

Power IFL takes advantage of the following aspects for clients willing to consolidate Linux workloads on the POWER Architecture:

- Competitive pricing to add Linux to an enterprise Power System
- Scalable to 32 sockets with seamless growth
- Enterprise-class reliability and serviceability

1.4 VIOS 2.2.3

The VIOS has been enhanced with flexibility, scalability, and resiliency features in the following areas:

- Shared Storage Pools
- Simplified Shared Ethernet Adapter
- Enhanced performance on Live Partition Mobility

1.5 VIOS Performance Advisor

The VIOS Performance Advisor tool provides advisory reports that are based on key performance metrics from various partition resources collected from the VIOS environment. This tool provides health reports that have proposals for making configurational changes to the VIOS environment and to identify areas to investigate further.

The VIOS Performance Advisor has been enhanced to provide support for N_Port Identifier Virtualization (NPIV) and Fibre Channel, Virtual Networking and Shared Ethernet Adapter, and Shared Storage Pool configurations.

1.6 PowerVM Live Partition Mobility

PowerVM Live Partition Mobility enhancements include system evacuation and a set of recommended settings to improve performance.

Server evacuation is a new feature that helps systems administrators to move all the capable logical partitions (LPARs) from one system to another when performing maintenance tasks and without disrupting business operations. This enhancement supports Linux, AIX, and IBM i VMs.

Partition mobility performance is improved by installing the latest available hardware firmware, Hardware Management Console (HMC), and VIOS software on both the source and target partitions.

Both enhancements are part of the Hardware Management Console V7.7.8 and no additional charge.

1.7 HMC feature updates

With release V7.7.8, the HMC is updated to include the following new functions:

- Support for Power Enterprise Pool management.
- IBM Power Virtualization Center Standard Edition enablement.
- User-defined thresholds to enable monitoring and alerting for workloads that can benefit from the Dynamic Workload Optimizer (DWO), as well as optional automation to invoke DWO when the threshold is exceeded. This function includes DWO to indicate whether a virtual machine will benefit from DWO.

- Additional tracking of dynamic logical partition activity within the current profile, which enables the reactivation of a virtual machine with all configuration changes intact since the last shutdown.
- Improved group-based access control for Lightweight Directory Access Protocol (LDAP) users, which enables limiting users to a subset of HMCs.

1.7.1 Considerations and prerequisites for HMC

The following list shows the minimum requirements and prerequisites to update the HMC and work with the new enhancements:

- The Power Enterprise Pools and DWO enhancements for HMC require the HMC at firmware level 7.7.8, or later
- To be able to use and manage Power Enterprise Pools or to use with IBM PowerVC, the HMC requires at least 2 gigabytes (GBs) of physical memory

The following HMC models cannot be upgraded to support this functionality and HMC V7.7.8 is their last supported firmware level:

- 7042-CR4
- 7310-CR4
- 7310-C05
- 7310-C06
- 7042-C06
- 7042-C07
- 7315-CR3
- 7310-CR3

The new graphical user interface (GUI) functionality is automatically disabled. The HMC operation then continues in legacy mode for HMC models with less than 2 GB of memory.

1.7.2 HMC and Power Enterprise Pool interaction

The HMC can be used to perform the following functions:

- The Mobile Capacity Upgrade on Demand (CoD) processor and memory resource activations can be assigned to systems with inactive resources. Mobile CoD activations remain on the system to which they are assigned until they are removed from the system.
- New systems can be added to the pool and existing systems can be removed from the pool.

- New resources can be added to the pool or existing resources can be removed from the pool.
- Pool information can be viewed, including pool resource assignments, compliance, and history logs.

1.7.3 HMC and IBM PowerVC interaction

IBM PowerVC manages PowerVM virtualization environments through a set of application programming interfaces (APIs) interacting with the HMC. These APIs provide to the HMC the necessary instructions to manage the Power Systems hardware, Power Hypervisor, and VIOS.

2

IBM Power Virtualization Performance

IBM Power Virtualization Performance (PowerVP) for Power Systems is a new product that offers a performance view into the PowerVM virtualized environment. It is capable of showing which virtual workloads are using specific physical resources on an IBM Power Systems server.

It is a solution that helps reduce time and complexity to find out and display performance bottlenecks. It presents to an administrator with a simple dashboard showing the performance health of the system. It can help simplify both prevention and troubleshooting and therefore reduce the cost of performance management.

It assists you in the following way:

- Shows workloads in real time, highlighting possible problems or bottlenecks (overcommitted resources)
- Helps to better use virtualized IBM Power System servers by showing the distribution of the workload
- Replays saved historical data
- Helps with the resolution of performance-related issues
- Helps to proactively address future issues that can affect performance

IBM PowerVP is integrated with the PowerVM Hypervisor and collects performance data directly from PowerVM Hypervisor, which offers the most accurate performance information about virtual machines (VMs) running on IBM Power Systems. This performance information is displayed on a real-time, continuous graphical user interface (GUI) dashboard, and it is also available for historical review.

IBM PowerVP offers these features:

- Real-time, continuous graphical monitor (dashboard) that delivers an easy-to-read display that shows the overall performance health of the Power server
- Customizable performance thresholds that enable you to customize the dashboard to match your monitoring requirements
- Historical statistics that enable you to go back in time and replay performance data sequences to find out performance bottlenecks
- System-level performance views that show all logical partitions (LPARs) and VMs and how they use real system resources
- VM drill-down (in-depth view), which gives you more performance details for each VM, displaying detailed information about various resources, such as CPU, memory, and disk activity
- Support for all VM types, including AIX, IBM i, and Linux
- Background data collection, which enables performance data to be collected when the GUI is not active

IBM PowerVP even allows an administrator to drill down to view specific adapter, bus, or CPU usage. An administrator can see the hardware adapters and how much workload is placed on them. IBM PowerVP provides both an overall and detailed view of IBM Power System server hardware so that it is easy to see how VMs are consuming resources.

2.1 Planning for an IBM PowerVP installation

IBM PowerVP is offered in a single standard edition. This edition is sold as a stand-alone additional offering for PowerVM Standard Edition clients and is included with PowerVM Enterprise Edition.

PowerVP has these prerequisites:

- IBM POWER7+™ and newer server models.
- IBM POWER7® C model servers.
- Firmware level 770, 780, or higher.

  For more details about supported machine model types, visit this website:

  https://www-304.ibm.com/webapp/set2/sas/f/power5cm/power7.html

  Note: At the time of writing this book, the firmware was supported on the following servers:

  - 8231-E1D (IBM Power 710 Express)
  - 8202-E4D (IBM Power 720 Express)
  - 8231-E2D (IBM Power 730 Express)
  - 8205-E6D (IBM Power 740 Express)
  - 8408-E8D (IBM Power 750)
  - 9109-RMD (IBM Power 760)
  - 9117-MMC (IBM Power 770)
  - 9179-MHC (IBM Power 780)
  - 8246-L1D (IBM PowerLinux 7R1)
  - 8246-L2D (IBM PowerLinux 7R2)
  - 8246-L1T (IBM PowerLinux 7R1)
  - 8246-L2T (IBM PowerLinux 7R2)
  - 8248-L4T (IBM PowerLinux 7R4)

- AIX operating system releases 6.1 and 7.1 are supported.
- VIOS versions are supported with the POWER7 and POWER7+ hardware.
- The Linux operating system releases, Red Hat Enterprise Linux (RHEL) 6.4 or later and SUSE Linux Enterprise Server 11 SP3 or later, are supported.
- The IBM i operating system Release 7.1 with Technology Refresh 6 (TR6) is supported. Program temporary fix (PTF) SI50350 for 5770SS1 is required. This PTF has prerequisites and corequisites that will be included when you order it.

- The GUI client is supported on platforms where Java Swing applications are supported. Installers are included for the following operating systems:
  - Windows
  - Linux
  - AIX
  - HPUX
  - Mac OS X
  - Solaris
  - UNIX
  - Basic Java

  IBM Java Runtime Environment (JRE) 1.6.0 has been tested and is supported.
- The GUI viewer can be installed in multiple clients, and multiple clients can connect and view data simultaneously. However, each connected client adds workload to the system while it requests performance data from the agents every second.
- IBM PowerVP needs at least one of the LPARs on each Power Systems server to be identified as a partition that will collect the system-wide information for all partitions. This LPAR is also known as the *system-level agent*. The other partitions are *partition-level agents*.

2.2 IBM PowerVP setup and usage

The IBM PowerVP agent is installed on the partitions on your POWER7 processor-based server. The partitions that you identify as the system-level agents need to be installed first. If you want to view partition-specific information for a partition, you also need to have the IBM PowerVP agent installed and running on those partitions (referred to as partition-level agents).

The system-level agent also acts as a partition-level agent for the partition on which it is running. The other partitions are then configured to point the partition-level agent to the system-level agent using the TCP/IP host name of the system-level agent partition. The partition-level agents need to connect to the system-level agent, so the system-level agent needs to be running before the partition-level agents can collect and provide partition-specific information. The system-level agent also needs to be running for the GUI to display information about the system and its partitions (Figure 2-1 on page 12).

The IBM PowerVP product installer is a graphical installer with a dialog. The PowerVP GUI is installed only on the system where you run the installation.

The agent installation is done automatically for IBM i using Restore licensed program (**RSTLICPGM**) with remote commands using the Java toolbox. The agent installation for AIX and VIOS is packaged in the **installp** format. The `installp` file set needs to be copied to the AIX partition to complete the installation. The agent installation for Linux is packaged in the Red Hat Package Manager (RPM) format. The RPM packages need to be copied to the Linux partition to complete the installation. File Transfer Protocol (FTP) or Secure Copy Protocol (SCP) can be used to move the installation files to AIX, VIOS, or Linux partitions.

2.2.1 Installation instructions

The installation of the various components is described.

Graphical user interface (GUI)

Figure 2-1 on page 12 shows how individual components of IBM PowerVP are connected together. The data flows in the following manner:

- The GUI reads data from the system agents on individual IBM Power System servers. There can be up to two system agents on a server.
- The system agents read (pull) data from the IBM POWER Hypervisor™.
- The system agents collect data from the partition agents and, on request, send them to the GUI.
- The partition agents push data to the system agents.

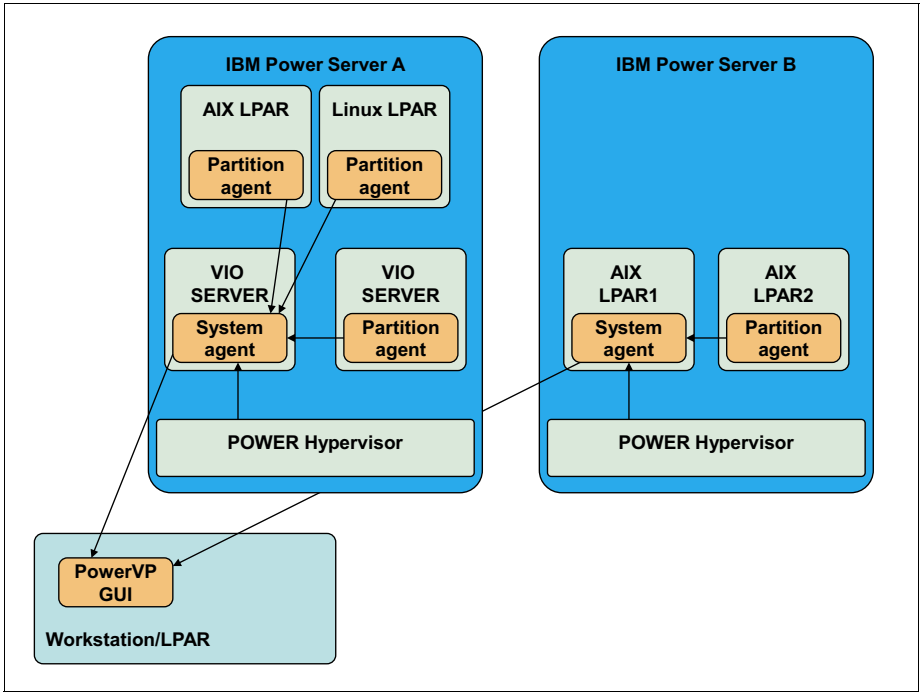


Figure 2-1 IBM PowerVP data flow

System agent

Follow these steps to get the system agent up and running:

1. Copy the installation package `powervp.x.x.x.x.bff` to a directory on the AIX and VIOS VM. From that directory, run the following commands as root:

   ```
   installp -agXd . powervp.rte
   cd /tmp/gsk8
   installp -acgqw -d /tmp/gsk8 GSKit*
   /opt/ibm/powervp/iconfig Listen="* 13000" SystemLevelAgent=
   ```

   The agent's configuration file is configured by the previous **`iconfig`** command, and it is in the following location:

   `/etc/opt/ibm/powervp/powervp.conf`

2. To start the IBM PowerVP partition agent for the first time without rebooting, run this command:

   `nohup /opt/ibm/powervp/PowerVP.sh &`

   After the next reboot, the agent is started automatically by the **`init`** script:

   `/etc/rc.d/rc2.d/SPowerVP`

Note: There is no need to run the IBM PowerVP partition agent on an LPAR with the system agent installed. The system agent also collects the same detailed statistics that the partition agent collects.

AIX partition agent

Follow these steps to get the AIX partition agent up and running:

1. Copy the installation package `powervp.x.x.x.x.bff` to a directory on the AIX and VIOS VM. From that directory, run the following commands as root:

   ```
   installp -agXd . powervp.rte
   cd /tmp/gsk8
   installp -acgqw -d /tmp/gsk8 GSKit*
   ```

2. To configure the partition agent to send data to the system agent (`vioa2` in our example), run this command:

   ```
   /opt/ibm/powervp/iconfig Listen="* 13000" SystemLevelAgent=vioa2
   ```

   The agent's configuration file is configured by the previous **`iconfig`** command, and it is in the following location:

   ```
   /etc/opt/ibm/powervp/powervp.conf
   ```

3. To start the PowerVP partition agent for the first time without rebooting, run this command:

   ```
   nohup /opt/ibm/powervp/PowerVP.sh &
   ```

   After the next reboot, the agent is started automatically by the **`init`** script:

   ```
   /etc/rc.d/rc2.d/SPowerVP
   ```

Linux partition agent

The prerequisites and the steps to get the Linux partition agent up and running are described.

These prerequisites need to be installed on the Linux system before the agent installation by running this command:

```
sysstat procps net-tools ethtool perf coreutils ksh
```

The following IBM PowerVP RPM files are needed on Linux systems:

- `gskcrypt64-8.0.50.11.linux.ppc.rpm`
- `gskssl64-8.0.50.11.linux.ppc.rpm`
- `powervp-driver-*.ppc64.rpm` (select the correct file that matches the Linux distribution installed)
- `powervp-x.x.x.x.ppc64.rpm`

Follow these steps to complete the installation and activation of the agent:

1. Copy all of the necessary RPM files to a directory on the Linux system. Follow these steps to list the installation requirements that are not met on the current system (the example shows that we used RHEL 6.4 Linux):

   ```
   rpm -qpR powervp-1.1.0-1.ppc64.rpm
   powervp-driver-rhel6.4-2.6.32-358.el6.ppc64.rpm
   gskcrypt64-8.0.50.11.linux.ppc.rpm gskssl64-8.0.50.11.linux.ppc.rpm
   ```

2. Install any dependencies that are not on the system, and then run the following command to install the IBM PowerVP agent (RHEL 6.4):

   ```
   rpm -i powervp-1.1.0-1.ppc64.rpm
   powervp-driver-rhel6.4-2.6.32-358.el6.ppc64.rpm
   gskcrypt64-8.0.50.11.linux.ppc.rpm gskssl64-8.0.50.11.linux.ppc.rpm
   ```

3. To configure the partition agent to send data to the system agent (`vioa2` in our example), run this command:

   ```
   /opt/ibm/powervp/iconfig Listen="* 13000" SystemLevelAgent=vioa2
   ```

   The agent's configuration file is configured by the previous **`iconfig`** command and it is in the following location:

   ```
   /etc/opt/ibm/powervp/powervp.conf
   ```

4. To start the IBM PowerVP partition agent for the first time without rebooting, run this command:

   ```
   nohup /opt/ibm/powervp/PowerVP.sh &
   ```

   After the next reboot, the agent is started automatically by the **`init`** script:

   ```
   /etc/rc.d/rc2.d/SPowerVP
   ```

Note: If there is no powervp-driver RPM that matches the version of Linux that is used, the source package `powervp-driver-source-1.1.0-1.ppc64.rpm` can be installed. This package installs the necessary source package to build a powervp-driver RPM on the current system. The files are unpacked in the `/opt/ibm/powervp/driver-source` directory. From that directory, issue the **`make`** command to build a powervp-driver RPM file for the current Linux system. There are many necessary prerequisite packages when you build the kernel modules. Consult the online documentation for the Linux prerequisites.

IBM i agent

The IBM i agent is installed automatically by the installation procedure of the IBM PowerVP GUI or it can be installed later by running the **`PowerVP_IBMi_Agent.exe`** agent installer. You need to configure the host name or IP address of the IBM i server and the system administrator's credentials to successfully install the IBM i agent. Also, the system agent's host name or IP address must be specified. Figure 2-2 and Figure 2-3 on page 16 show examples of how an IBM i agent can be configured from the GUI installer.

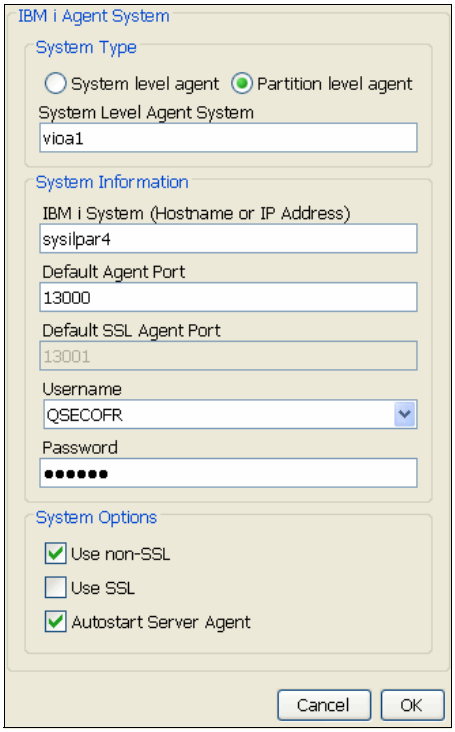


Figure 2-2 IBM PowerVP IBM i agent definition

Figure 2-3 IBM PowerVP IBM i agent ready to install

Use these commands to check whether the IBM i agent is successfully installed (Example 2-1).

Example 2-1 IBM i agent installation check

```
GO LICPGM
Select option 10
Check the list for product 5765SLE
```

If you need to install the IBM i agent manually, use the following instructions in Example 2-2.

Example 2-2 IBM i manual agent installation

```
Create a new savf on the IBM i system using:
CRTSAVF library/filename
FTP qsle.savf in binary mode to the savf created on IBM i system using:
put qsle.sav[library]/[filename]
Run:
RSTLICPGM LICPGM(5765SLE) DEV(*SAVF) SAVF(library/filename)
CALL QSLE/QPFICONFIG PARM('config-keyword=config-value'
'config-keyword=config-value')
```

2.2.2 IBM PowerVP use

Before your first use, understand the various functions of the GUI.

The sections of the IBM PowerVP GUI main window are shown in Figure 2-4 and described:

- The System Information section provides information for the POWER7 system on which the system-level agent runs.
- The System Usage section provides system-wide performance information, including a running graph of total CPU utilization.
- The Playback section provides the interface to record into a file the information that the IBM PowerVP GUI is displaying.
- The Host Information section provides the systems to which the GUI is connected and from which performance data is received.

System-wide statistics

The partition list at the top contains a line for every LPAR on the Power Systems server. The first column indicates the partitions that can be “drilled down” to see partition-specific performance information. The second column is the LPAR ID, which matches the configuration in the HMC for the system. The third column indicates whether the processors for the partition are Dedicated or Shared. The fourth and fifth columns provide the Cores Entitled and Cores Assigned (currently using) for the partition. The sixth column is a moving bar that indicates the CPU utilization for the partition. An example of the IBM PowerVP main window is in Figure 2-4.

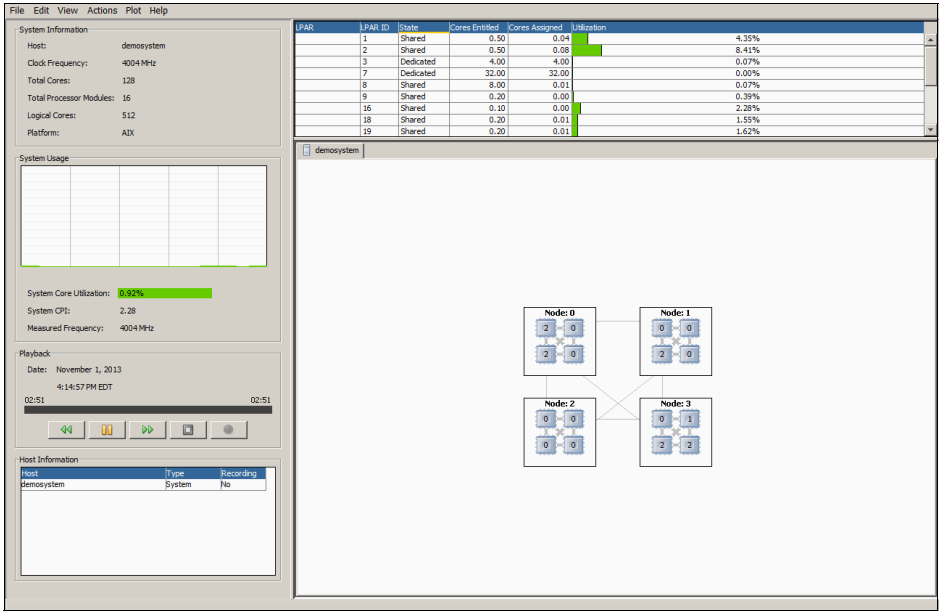


Figure 2-4 IBM PowerVP Dashboard main window

The largest section is a graphical representation of Power Systems server processors and buses. You see a box for each node on your system. Inside the node boxes are smaller shaded boxes for each processor module on the node. The lines between the processor modules are for the buses that connect the processor modules to each other within the node. The lines between the nodes are for the buses that connect the nodes to each other. The colors you see indicate the level of utilization of the bus.

A new tab is displayed after you select a node, showing the in-depth (or a *drill-down* view) hardware of the selected node. The larger boxes are the processor modules within the node. Columns are in each processor module box that indicate each of the CPU cores on the module. The utilization depicted in the cores will change over time as performance statistics change, and possibly the color will also change. The lines between the processor modules represent the buses between the modules. The lines that run off the page represent the buses to other nodes.

The boxes above and below the processor modules represent the I/O controllers (also known as the *GX controllers*) with the lines to them representing the buses from the processor modules to the controllers. Similarly, the boxes to both sides represent the memory controllers (also known as the *MC controllers*) with the lines to them representing the buses from the processor modules to the controller. The colors of the lines can change based on the utilization of the buses. The bus utilization is also shown as a percentage in the controller box.

For partitions with dedicated cores, you can click the LPAR line to show the cores that are assigned to the partition. You can also click any of the cores to show which LPAR is assigned to the core. If an LPAR or core is assigned to a shared partition pool, these are all grouped together with the same color (usually blue) because they cannot be differentiated. If you have active cores that are not assigned to a dedicated partition or the shared pool, these cores can have CPU utilization because they might be borrowed by partitions that need additional processing power.

LPAR statistics

To select an individual LPAR, double-click one of the host names in the column. An example of IBM PowerVP LPAR drill-down (in-depth view) is in Figure 2-5.

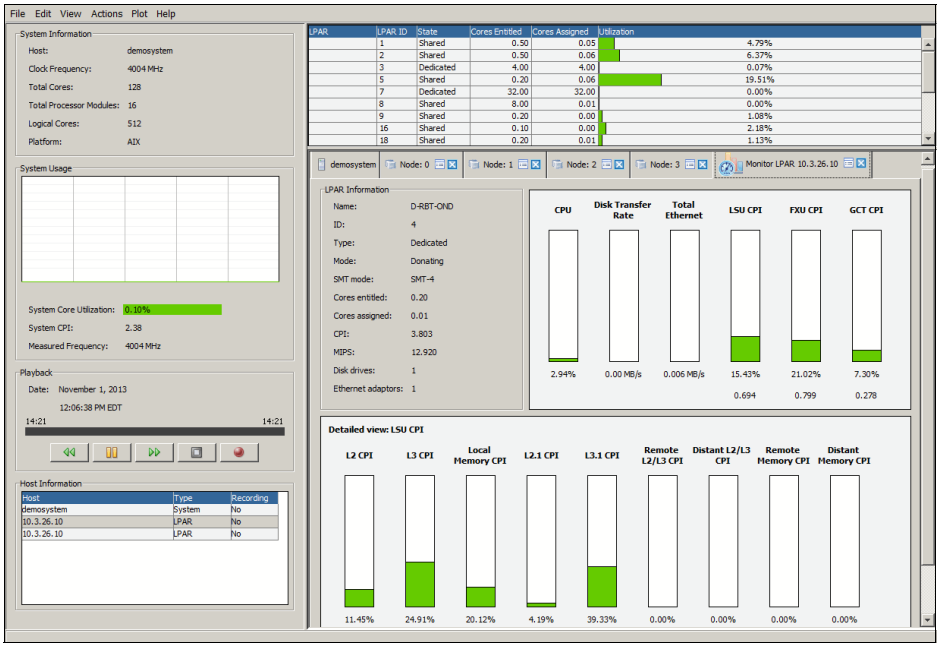


Figure 2-5 PowerVP Dashboard partition details

If you drill down to a specific LPAR, the following information is displayed. A new tab is created that shows the partition detailed information. The bars represent different performance metrics for that specific partition:

- The CPU column shows the CPU utilization for that partition as a percentage of the entitled processor resources.
- The Disk Transfer Rate shows the rate of bytes read and written to disk. After selecting a disk column, statistics for the individual disks for the LPAR in the bottom half of the display will appear.
- The Total Ethernet column represents the rate of bytes sent and received on the Ethernet. After selecting the network column, statistics for each Ethernet adapter for the LPAR in the bottom half of the display will appear.

- The LSU CPI represents the cycles spent on Load/Store Unit resources for the partition. This includes the whole memory hierarchy from local caches to distant memory. If you click this column, you will see a breakdown of the LSU CPI for the LPAR in the bottom half of the display. This information is retrieved from the Performance Monitoring Unit (PMU) in the POWER hardware using APIs on AIX, VIOS, and Linux, and using Performance Explorer (PEX) on IBM i. On IBM i, a 30-second PEX collection is used; therefore, the columns update every 30 seconds. On AIX, VIOS, and Linux, the collection interval is shorter; therefore, the columns are updated more frequently.
- The FXU CPI represents the cycles spent on Floating Point execution for the partition.
- The GCT CPI represents the cycles spent waiting on the Global Completion Table (GCT) for the partition. The Global Completion Table is used for pipelining Out Of Order execution.

2.2.3 Hands-on example demonstration with screen captures

The example screen captures in this section were taken on the following demonstration environment:

- IBM PowerVP GUI running on a Microsoft Windows 7 workstation
- IBM PowerVP system agent running on AIX 7.1 TL2SP2 - LPAR (ID 4). LPAR ID 4 was enabled to collect system statistics by the HMC administrator.
- IBM PowerVP partition agent running on AIX 7.1 TLSP2 - LPAR ID 5.
- The server hardware was IBM Power 795 (9119-FHB) with the following features:
  - Firmware Release AH780_028 (firmware version 770 or 780 is a prerequisite of IBM PowerVP).
  - There were 128 CPU cores in the machine. The cores were installed in four books (nodes). Each book has four Single Chip Modules (SCMs). Each SCM is equipped with eight POWER7 cores running at 4 GHz.
  - Two TBs of memory are placed in four books and distributed over memory banks.

When you make a new connection from IBM PowerVP GUI to a system agent and the connection is established, you immediately see the window presented in Figure 2-6. The dashboard is divided into several areas:

System information Shows basic information about server hardware

System usage Shows aggregated system usage (CPU) for all running LPARs

Playback Displays record functions and playback functions

System information Shows the name of the server to which you are connected through a system agent (`demosystem`)

Individual LPARs Shows each LPAR on a single line. There are many LPARs in our demonstration machine.

Graphical view of the system
Shows that the system is configured with four processor books (nodes) in our demonstration

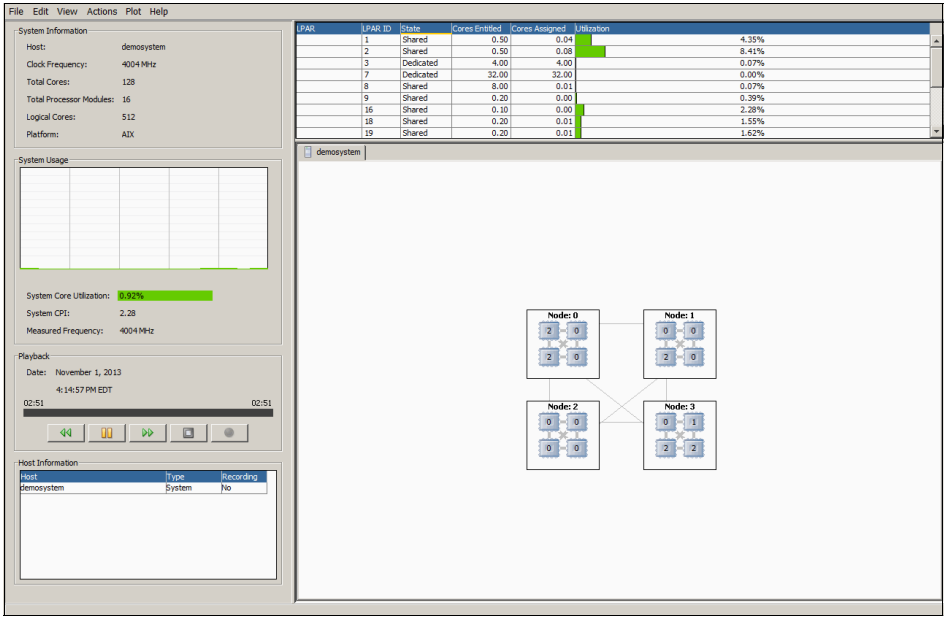


Figure 2-6 GUI Dashboard - Main window

After clicking one of the nodes in the main graphical panel of the dashboard, IBM PowerVP GUI drills down. You see the detailed information about the hardware and a graphical representation of how heavily the individual components of the node are used. In Figure 2-7, we can see a drill-down (in-depth view) into node 1 and the utilization level of its components. You can see the following information:

- Only one SCM module, module 2, is lightly used (CPU cores 0 and 7).
- Other SCMs are not used at the moment of capturing the image.
- Two memory banks of module 2 are slightly used.
- The bus between module 2 and module 0 is slightly used.

Figure 2-7 IBM PowerVP dashboard - Node 1 utilization

For LPARs running in dedicated mode, it is possible to see which CPUs are assigned to that LPAR. Click a specific LPAR line in the upper-right section of the dashboard and search through nodes to see which CPUs belong to that LPAR. In our example in Figure 2-8, LPAR ID 7 uses all CPU cores from Node 2.

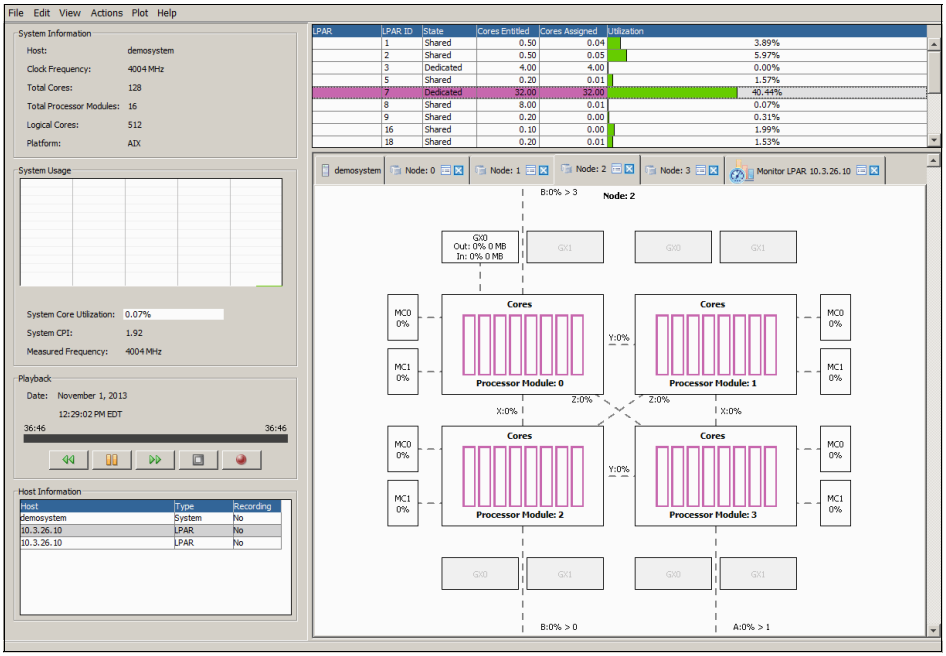


Figure 2-8 IBM PowerVP dashboard - Dedicated CPU LPAR

If you double-click an LPAR in shared CPU mode, all shared CPU LPARs in the same shared pool are selected and you can see which CPUs in which nodes belong to that shared processor pool. In Figure 2-9, the shared processor pool is used by many LPARs and it is used by the Hypervisor on all cores from CPU Module 2 (cores 0 - 7) of Node 1.

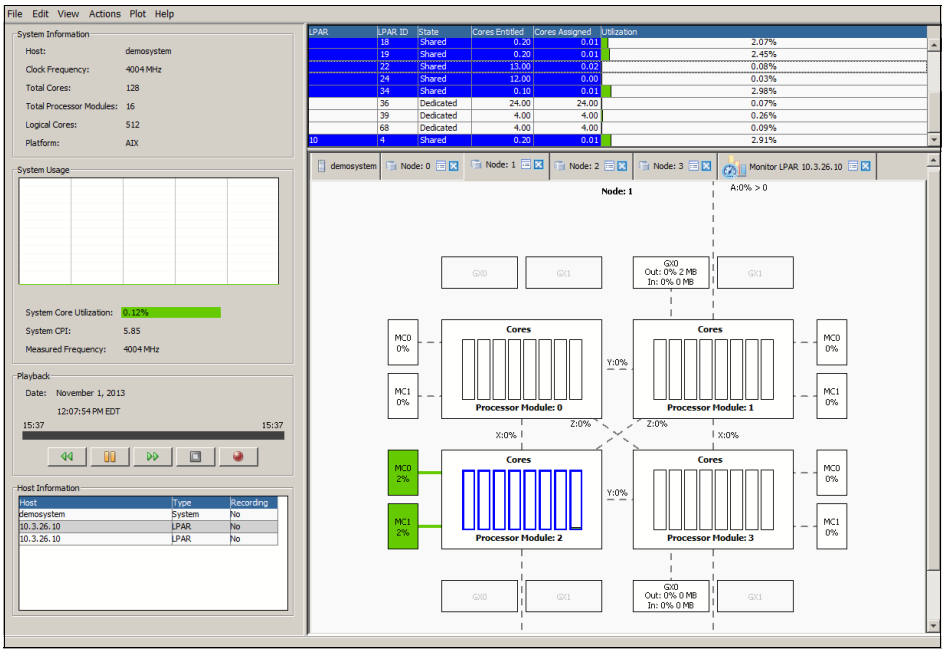


Figure 2-9 IBM PowerVP Dashboard - Shared CPU pool

For LPARs with a partition agent installed, you can drill down to the LPAR's statistics by double-clicking a specific LPAR in the upper-right area of the dashboard. As shown in Figure 2-10, new statistics appear. In our example, the system agent runs on LPAR ID 4 and partition agents run on LPAR IDs 4 and 5. Figure 2-10 shows detailed statistics about the CPU of LPAR ID 4 (our system agent).

Note: The IBM PowerVP system agent also behaves like a partition agent. There is no need to run both system and partition agents on a single LPAR. Only LPARs that do not run system agents need partition agents.

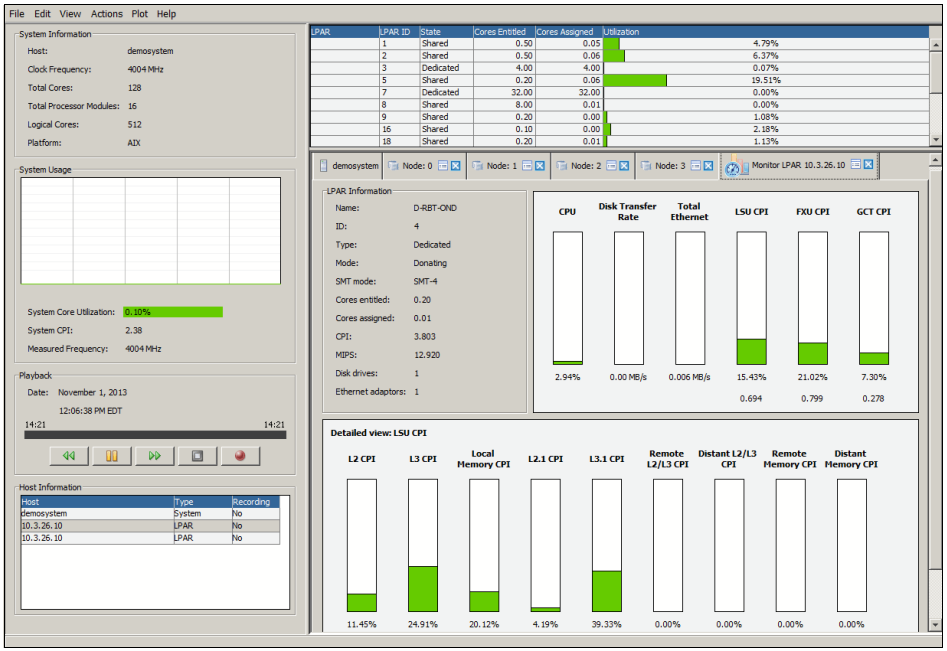


Figure 2-10 PowerVP dashboard - LPAR CPU drill down

It is also possible to display individual Ethernet adapter statistics as shown in Figure 2-11.

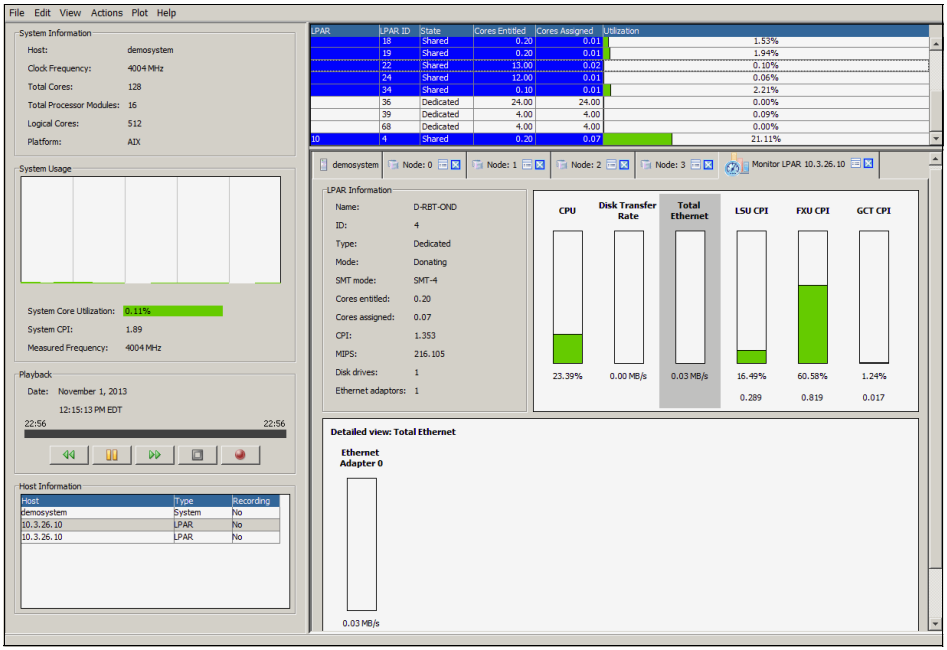


Figure 2-11 IBM PowerVP dashboard - LPAR Ethernet statistics

It is also possible to display individual physical disk statistics inside an LPAR with a partition agent as shown in Figure 2-12.

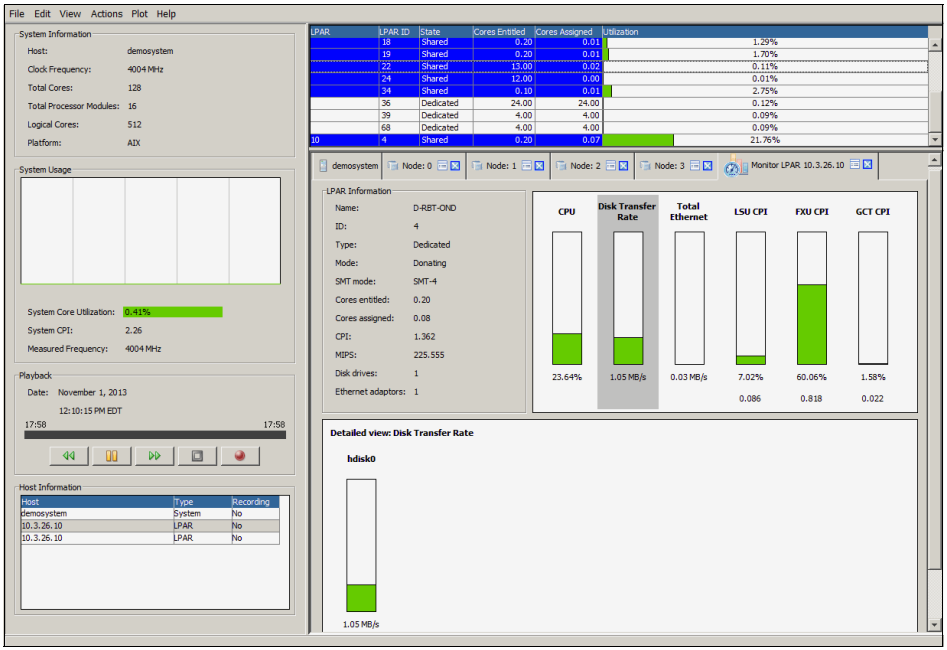


Figure 2-12 IBM PowerVP dashboard - LPAR hdisk statistics

3

Power Integrated Facility for Linux

Since 2000, IBM has continually invested in Linux on Power. With new initiatives, such as PowerLinux Centers worldwide (Austin, TX; Beijing, China; Montpellier, France; and New York, NY), IBM maintains its strong engagement on this industry standard.

Power Integrated Facility for Linux is a flexible and affordable high-performance capacity offering for Linux workloads.

With the Power Integrated Facility for Linux (Power IFL) offering, IBM brings the industry-leading class Power platform closer to the Linux ecosystem. Power IFL helps clients to consolidate operations and reduce overhead by using their existing production systems and infrastructures.

Power IFL is available for Power 770, 780, and 795 servers with available capacity on demand (CoD) memory and cores.

3.1 Structure and fulfillment

The Power IFL offering is meant to be simple and flexible. Here is a list of the details for fulfillment:

- Each Power IFL feature delivers four processor and 32 GB memory activations.
- Power IFL does not provide physical hardware (processor cards, books, or nodes).
- The PowerVM for PowerLinux license is entitled for the Power IFL cores on Power 770, 780, and 795 servers.
- The PowerVM for PowerLinux license entitlement and corresponding Software Maintenance agreement (SWMA) can coexist with a PowerVM Enterprise edition (EE) (for AIX and IBM i) license and an SWMA on a single system.
- Power clients agree to segregate Power IFL cores in a separate virtual shared processor pool from cores purchased to support AIX and IBM i.

Support is available through capacity on demand (CoD) activations. You can enable additional cores for Linux and Virtual I/O Server (VIOS) partitions on select Power Systems servers by ordering Feature Code (FC) ELJ0. In the following sections, we describe the requirements and supported systems.

3.1.1 Requirements

Power IFL has the following requirements:

- Firmware level 780
- HMC Level Version 7, Release 7.8

3.1.2 Supported systems

Activations are available on Power 770, Power 780, and Power 795 servers with the following models and types:

- 9119-FHB
- 9117-MMB
- 9179-MHB
- 9117-MMC
- 9179-MHC
- 9117-MMD
- 9179-MHD

Ordering the FC ELJ0 Power IFL package automatically enables the following feature codes:

- FC ELJ1 and FC ELJ4 = Four core activations
- FC ELJ2 = 32 GB memory activations
- FC ELJ3 = Four PowerVM for Linux entitlements

The Power IFL contract (form Z126-6230) must be signed by the client before the order. This contract needs to be signed one time for each client enterprise per country. The client agrees to run the system in a manner that isolates the Power IFL cores in a separate virtual shared processor pool from the rest of the other operating system cores.

Figure 3-1 shows a simple overview of the offering.

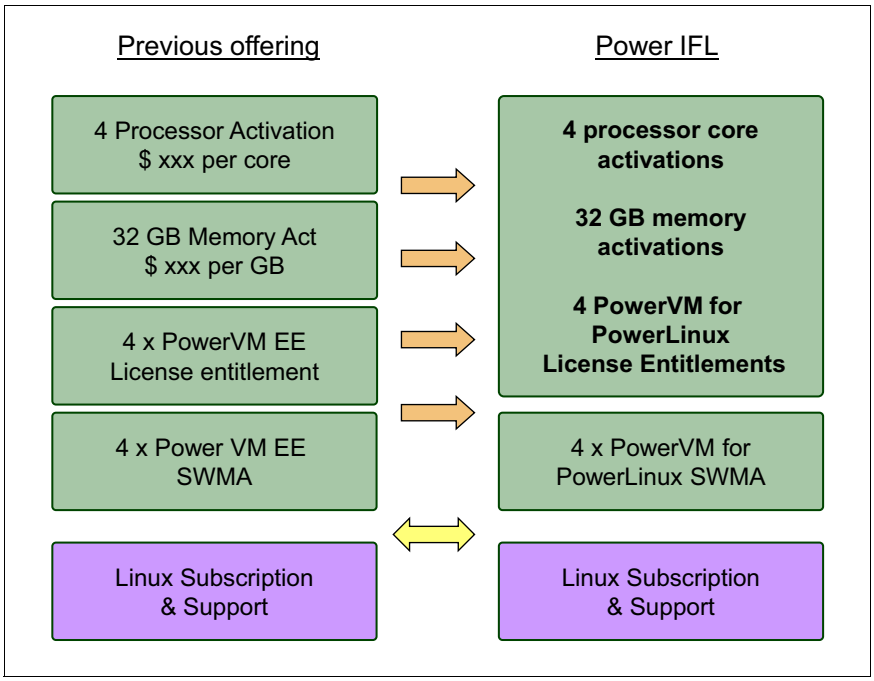


Figure 3-1 Power IFL simplified offering

Note: If PowerVM Standard Edition is running on other cores, all cores will be upgraded to PowerVM Enterprise Edition (5765-PVE) at the client's expense.

3.2 Configuration

The processing capacity of an IFL core must be used by Linux partitions only. This capacity cannot be used to satisfy the licensed core requirement for the VIOS, AIX, and IBM i partitions.

The number of general-purpose cores, and therefore the capability available for AIX and IBM i partitions, is the total number of licensed activations minus any IFL and VIOS activations.

It is possible to create an environment from the Hardware Management Console (HMC) to ensure that a system is kept in compliance with these license agreements. This can be achieved with a shared processor pool configuration. Shared processor pools can be configured on the HMC as shown in Figure 3-2.

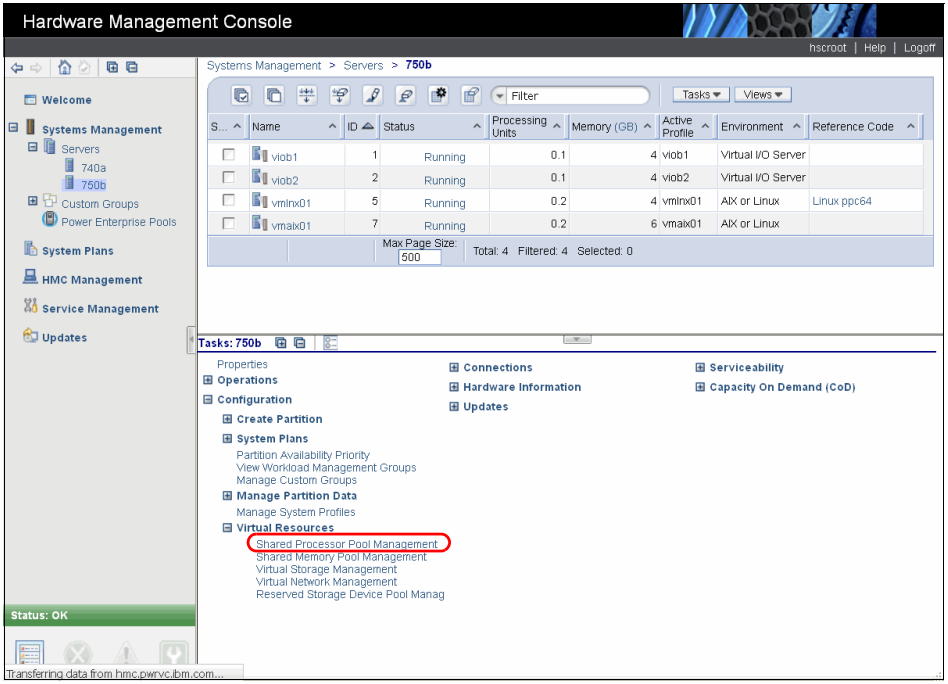


Figure 3-2 HMC Shared Processor Pool Management

Note: If an IFL core is enabled, this processing capacity must be used by Linux partitions and cannot be used to satisfy the licensed core requirement for VIOS, AIX, and IBM i partitions. Linux partitions can consume the capacity provided by general-purpose cores, VIOS cores, and IFL cores.

3.2.1 Compliance configuration examples

In this section, we demonstrate two configuration examples. More configuration examples and details are provided in the Power Systems Information Center under the IBM POWER7 Systems™ section:

http://pic.dhe.ibm.com/infocenter/powersys/v3r1m5

Power IFL activation scenario 1

Activation scenario 1 in Figure 3-3 shows a default of compliance according to the Power IFL statement.

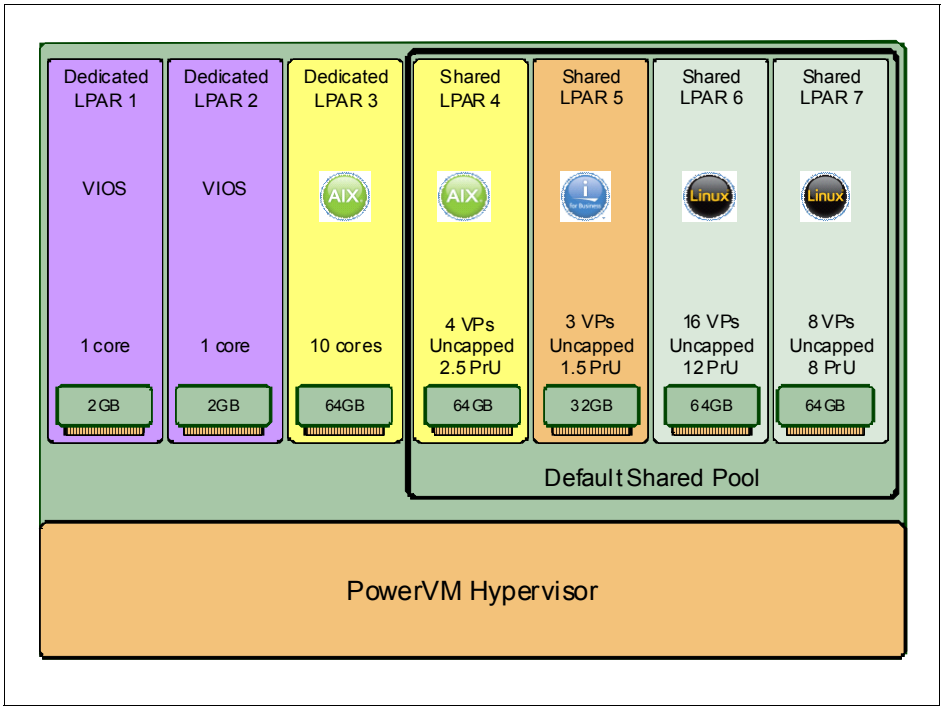


Figure 3-3 Power IFL activation scenario 1

The following points relate to scenario 1:

- Four Power IFLs are assigned to the default shared processor pool (16 cores and 128 GB activations).
- The Power IFL cores are being added to the shared processor pool and create two additional Linux partitions:
  - Linux LPAR 6 has 16 uncapped virtual processors with 12 processing units of entitlement.
  - Linux LPAR 7 has eight uncapped virtual processors with eight processing units of entitlement.

This configuration is out of compliance for its following aspects:

- The AIX and IBM i resource consumption exceeds the total number (16) of the licensed cores:
  - The total capacity available to the shared processor pool is seven processors.
  - Because the AIX shared partition is uncapped, it can consume up to four physical cores of processing capacity.
  - The IBM i is uncapped and can get up to three physical cores of processing capacity.
  - The VIOS and AIX dedicated partitions equal 12 processor cores.
- A total of 19 processor cores are available to AIX and IBM i.

Power IFL activation scenario 2

Figure 3-4 on page 35 shows an example method to solve the scenario 2 default of compliance.

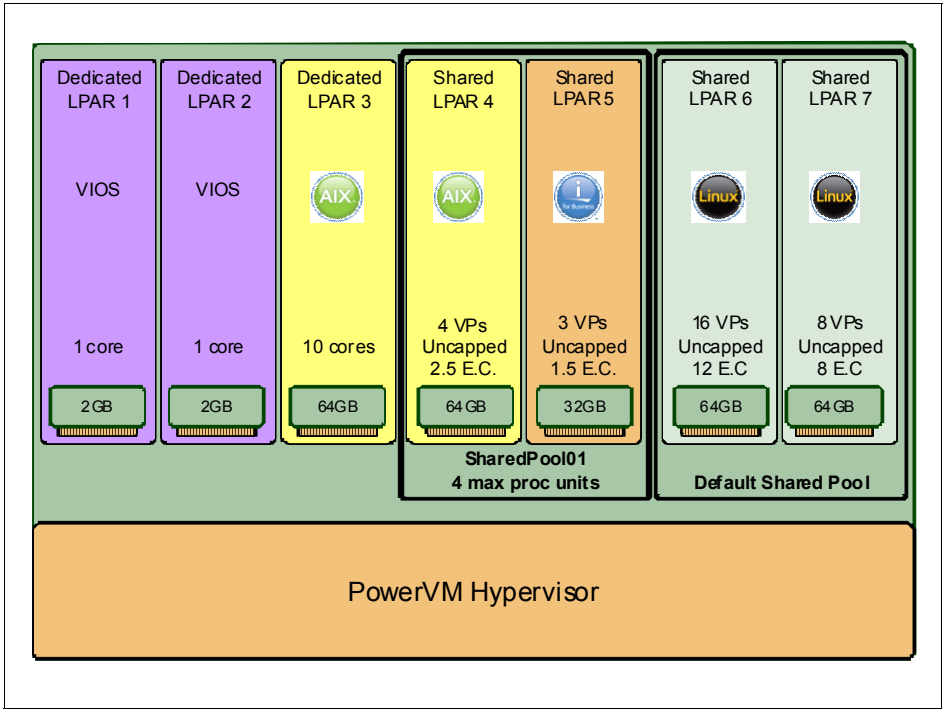


Figure 3-4 Power IFL activation scenario 2

To solve the compliance issues from scenario 1, the following actions are in place:

- The VIOS and dedicated AIX partition remains the same.
- The Maximum Processing Units for Shared Pool01 is set to 4 by way of the Virtual Resources section on the HMC to prevent AIX and IBM i from obtaining more than four physical processor cores of resources.
- The AIX and IBM i Shared LPARS are dynamically moved into the shared processor pool SharedPool01 by way of the Partitions tab on the Shared Processor Pool Management panel on the HMC.
- AIX, IBM i, and VIOS LPARs can only obtain a maximum of 16 cores.

3.2.2 Compliance Monitoring Assistance

The latest available Firmware 780 Service Pack 1 HMC version provides notification through the HMC when the system is out of compliance with the license agreement for the usage of IFL and VIOS core capacity.

The firmware periodically computes the possible maximum core consumption of the various license types (AIX, IBM i, and VIOS). If a system is out of compliance, a message panel is shown by the HMC to indicate that the system is not in compliance with the license agreement.

It will be the system owner's responsibility to bring the configuration back into compliance.

Note: At the time of writing this book, the firmware performs a soft compliance validation.

4

Virtual I/O Server 2.2.3

This chapter describes new enhancements made to the Virtual I/O Server (VIOS) in 2013. Shared Storage Pools (SSPs) have been enhanced to improve flexibility, scalability, and resiliency. These improvements include mirroring of the storage pool, dynamic contraction of the storage pool, dynamic disk growth within the storage pool, and scaling improvements. VIOS Performance Advisor has been enhanced to provide support for N_Port Identifier Virtualization (NPIV) and Fibre Channel, Virtual Networking and Shared Ethernet Adapter, and SSP configurations. The Shared Ethernet Adapter Failover configuration has been simplified by removing the extra complexity when configuring a control channel. Also, VIOS helps improve the Live Partition Mobility performance enhancement.

The main enhancements in 2013 are in the following areas:

- Simplified Shared Ethernet Adapter failover configuration setup
- SSP enhancements
- Shared Ethernet Adapter now, by default, uses the `largesend` attribute
- VIOS Performance Advisor enhancements (Chapter 5, “Virtual I/O Server Performance Advisor” on page 79)
- Support for Live Partition Mobility performance enhancements (Chapter 6, “PowerVM Live Partition Mobility” on page 95)

All the new 2013 features are included in VIOS Version 2.2.3. There are two ways to update the level:

- Fresh installation from media (DVD)
- Upgrade from the previous version (**updateios** command). The update packages can be downloaded from the IBM Fix Central website:

  http://www-933.ibm.com/support/fixcentral

From now on, the VIOS rootvg requires at least 30 GB of disk space. It is advised that you protect the VIOS rootvg by a Logical Volume Manager (LVM) mirror or hardware RAID. In correctly configured redundant VIOS environments, it is possible to update VIOSs in sequence without interrupting client virtual machines (VMs). Extra work might be required if client VMs are configured to use an LVM mirror between logical unit numbers (LUNs) that are provided by dual VIOSs.

4.1 Simplified SEA failover configuration

The Shared Ethernet Adapter (SEA) is the PowerVM component used to bridge the virtual networking to the physical network (physical network interface card). The SEA allows the client LPARs with virtual adapters to share the physical network resources with other LPARs and access the external network.

The PowerVM implementation of virtual networking takes place in both the Power Hypervisor and VIOSs.

In this section, we describe the new method used for SEA failover configuration. This enhancement is achieved by removing the requirement of a dedicated control-channel adapter for each SEA configuration pair.

4.1.1 Requirements

The new simplified SEA failover configuration is dependent on the following requirements:

- VIOS Version 2.2.3
- Hardware Management Console (HMC) 7.7.8
- Firmware Level 780 or higher

Note: At the time of writing this book, this feature is not supported on hardware models MMB and MHB.

For more details about the supported machine model types, go to this website:

https://www-304.ibm.com/webapp/set2/sas/f/power5cm/power7.html

4.1.2 Design overview

The SEA was enhanced to make it easier to configure. The requirement has been removed for a dedicated control-channel Ethernet adapter and a VLAN ID for each SEA failover configuration.

The SEA failover still supports the traditional provisioning of the dedicated control-channel adapter in SEA failover VIOSs. Existing SEA and SEA failover functionality continues to work, which allows the existing SEA failover configuration to migrate to the new VIOS. The new mechanism is supported without making any configuration changes.

By removing the requirement for a dedicated control channel for SEA pairs, the SEA implements a new method to discover SEA pair partners using the VLAN ID 4095 in its virtual switch. After partners are identified, a new SEA high availability (HA) protocol is used to communicate between them.

Multiple SEA pairs are allowed to share the VLAN ID 4095 within the same virtual switch. We still can have only two VIOSs for each SEA failover configuration.

The new simplified SEA failover configuration relies the following dependencies:

- VLAN ID 4095 is a reserved VLAN for internal management traffic. POWER Hypervisor 7.8 and higher have support for management VLAN ID 4095.
- The HMC ensures that the management VLAN ID 4095 is not user configurable.
- The HMC also needs to ensure that the SEA priority value is either 1 or 2 so that users do not configure more than two SEAs in a failover configuration.

Because the existing SEA failover configuration is still available, the following method is used to identify a simplified configuration:

- The method to discover an SEA failover partner is decided based on user input for the control channel (**`ctl_chan`**) attribute of the SEA device on the **`mkvdev`** command.
- If the control-channel adapter is specified on the **`mkvdev`** command and the specified adapter is not one of the trunk adapters of the SEA, a dedicated control-channel adapter is specified.
- If no control-channel adapter is specified on the **`mkvdev`** command, the default trunk adapter is the Port Virtual LAN Identifier (PVID) adapter of the SEA. Partners are discovered using the new discovery protocol implementation over the management VLAN ID 4095.

Important: The new implementation attempts to avoid miscommunication. Nevertheless, it is not able to prevent users from configuring one SEA with a dedicated control-channel adapter and the other SEA without the dedicated control-channel adapter in the HA configuration.

The administrator must ensure that a consistent configuration approach is used on both SEAs.

4.1.3 Usage considerations

To validate the simplified SEA failover partners' configuration, use the following steps:

- The SEA LPAR partners are identified by first comparing their PVIDs and then comparing the additional VLANs they bridge. The PVID must be unique to each SEA pair associated in the SEA failover configuration.
- If the PVID of the SEA in two different VIOSs matches, they are considered partners. They are still considered partners if their PVIDs match but their additional VLANs do not match. However, this is considered a transient condition and an error is logged for the mismatch of the VLANs.
- The SEA failover pairs need to have identical sets of VLANs at all times. It is possible to modify VLANs via the HMC (add or remove). This can create mismatched VLANs between SEA LPARs. This needs to be a transient condition only, and an error is logged in the VIOSs' errlogs.
- If the SEA pair is configured with matching VLANs but their PVIDs are different due to miscommunication, they are not considered SEA failover partners even though they have a matching VLAN set. This indicates two different VLAN IDs for untagged network traffic, which is a misconfiguration.

4.1.4 Migration

Migration from the current SEA to the simplified SEA configuration without a dedicated control channel requires a network outage. It is not possible to remove the dedicated control-channel adapter dynamically at run time.

The SEA must be in a defined state before you can remove the dedicated control-channel adapter. This is necessary to avoid any condition that leads to an SEA flip-flop or both SEAs bridging.

An update of the firmware to level 7.8 or a higher version is required to take advantage of the new discovery protocol. This firmware update from a lower service release is a disruptive update and requires a network outage.

4.1.5 Examples

The new syntax for the **mkvdev** command authorizes you to use the **ha_mode** parameter without specifying any control-channel adapter.

We do not have to specify any extra control-channel adapter, and the SEA configuration uses the new protocol to identify its partner with the VLAN 4095. The new command syntax is shown:

```
mkvdev -sea TargetDevice -vadapter VirtualEthernetAdapter -default
DefaultVirtualEthernetAdapter -defaultid SEADefaultPVID -attr
ha_mode=auto
```

Figure 4-1 is an example of a simplified SEA configuration.

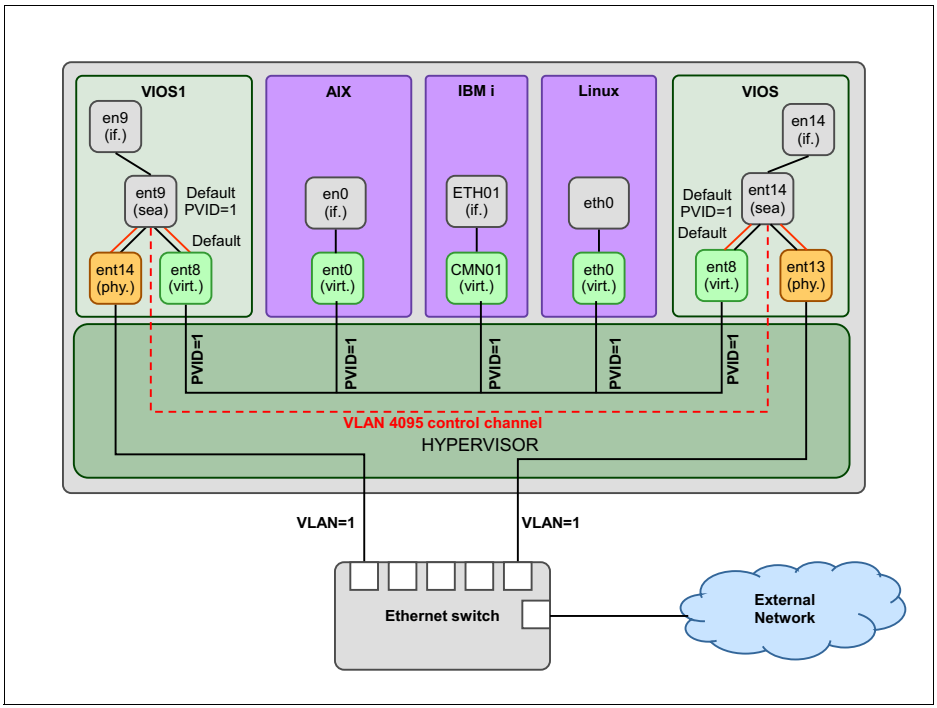


Figure 4-1 SEA simplified configuration example

Follow these steps:

1. Add the virtual Ethernet adapter with PVID 144 with a slot number 144 to VIOS1 from the HMC as shown in Figure 4-2 by using a dynamic logical partitioning operation.

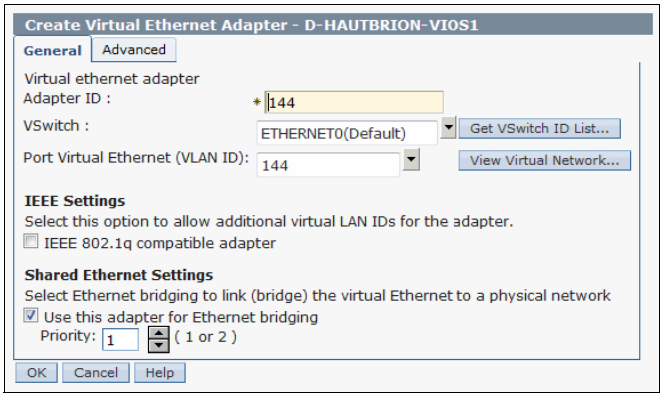


Figure 4-2 Add virtual Ethernet adapter to VIOS1

2. We must identify the new virtual adapter from VIOS1 with the **lsdev** command. We create the SEA adapter with the **mkvdev** command using the physical Ethernet adapter ent14. The SEA adapter ent9 is created:

```
$ lsdev -dev ent* -vpd|grep C144|grep ent
  ent8            U9119.FHB.5102806-V1-C144-T1  Virtual I/O
Ethernet Adapter (l-lan)
$ mkvdev -sea ent14 -vadapter ent8 -default ent8 -defaultid 144
-attr ha_mode=auto
ent9 Available
en9
et9
```

3. We verify that ent9 is now the primary adapter with the **errlog** command:

```
$ errlog
IDENTIFIER TIMESTAMP  T C RESOURCE_NAME  DESCRIPTION
E48A73A4   1104235213 I H ent9           BECOME PRIMARY
```

4. Add the virtual Ethernet adapter with PVID 144 with a slot number 144 to VIOS2 from the HMC as shown in Figure 4-2 on page 43 using a dynamic logical partitioning operation.

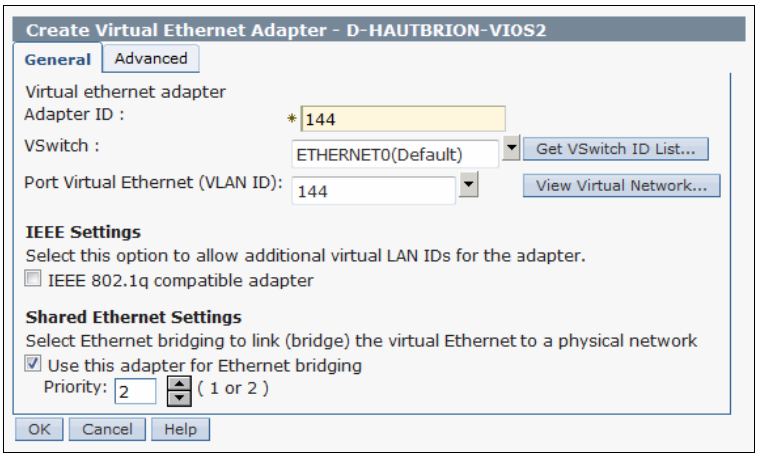


Figure 4-3 Add virtual Ethernet adapter to VIOS2

5. We must identify the new virtual adapter from VIOS2 with the **lsdev** command. We create the SEA adapter with the **mkvdev** command using the physical Ethernet adapter ent8. The SEA adapter ent14 is created:

```
$ lsdev -dev ent* -vpd|grep C144|grep ent
  ent13            U9119.FHB.5102806-V2-C144-T1  Virtual I/O
Ethernet Adapter (l-lan)
$  mkvdev -sea ent8 -vadapter ent13 -default ent13 -defaultid 144
-attr ha_mode=auto
ent14 Available
en14
et14
```

6. We check that ent14 is now the backup adapter with the **errlog** command:

```
$ errlog
IDENTIFIER TIMESTAMP  T C RESOURCE_NAME  DESCRIPTION
1FE2DD91   1104165413 I H ent14          BECOME BACKUP
```

7. We verify that VIOS2 can fail over by putting the VIOS1's SEA adapter in standby mode with the **chdev** command and check the errlog entry:

```
chdev -dev ent9 -attr ha_mode=standby
$ errlog
IDENTIFIER TIMESTAMP  T C RESOURCE_NAME  DESCRIPTION
1FE2DD91   1104235513 I H ent9           BECOME BACKUP
```

8. The VIOS2's SEA adapter is now the primary adapter:

```
$ errlog
IDENTIFIER TIMESTAMP  T C RESOURCE_NAME  DESCRIPTION
E48A73A4   1104165513 I H ent14          BECOME PRIMARY
```

9. We also check the current primary and backup adapters with the **entstat** command on both VIOSs:

   – VIOS1:

```
$ entstat -all ent9|grep Active
  Priority: 1  Active: False
  Priority: 1  Active: False
```

   – VIOS2:

```
$ entstat -all ent14|grep Active
  Priority: 2  Active: True
  Priority: 2  Active: True
```

10. We return to the initial state with the **chdev** command one more time:

```
chdev -dev ent9 -attr ha_mode=auto
IDENTIFIER TIMESTAMP  T C RESOURCE_NAME  DESCRIPTION
E48A73A4   1104235913 I H ent9           BECOME PRIMARY
```

11. We monitor the SEA failover state from the **entstat** command and search for the "High Availability Statistics" stanza. We confirm that this SEA is using VLAN ID 4095 as its control channel and verify the current primary bridging adapter:

   – VIOS1:

```
$ entstat -all ent9
../..
High Availability Statistics:
    Control Channel PVID: 4095
    Control Packets in: 1306
    Control Packets out: 1121
Type of Packets Received:
    Keep-Alive Packets: 910
    Recovery Packets: 1
    Notify Packets: 1
    Limbo Packets: 0
    State: PRIMARY
    Bridge Mode: All
    Number of Times Server became Backup: 1
    Number of Times Server became Primary: 2
    High Availability Mode: Auto
    Priority: 1
```

– VIOS2:

```
$ entstat -all ent14
../..
High Availability Statistics:
    Control Channel PVID: 4095
    Control Packets in: 887
    Control Packets out: 1333
Type of Packets Received:
    Keep-Alive Packets: 466
    Recovery Packets: 1
    Notify Packets: 1
    Limbo Packets: 0
    State: BACKUP
    Bridge Mode: None
    Number of Times Server became Backup: 2
    Number of Times Server became Primary: 1
    High Availability Mode: Auto
    Priority: 2
```

4.2 SSP enhancements

The Shared Storage Pool (SSP) functionality included in VIOS is constantly enhanced. VIOS Version 2.2.3 includes the latest (2013) improvements, which are also referred to as *Release 4 of SSP*.

In the latest release of VIOS Version 2.2.3, the SSP functionality has been further enhanced in the following areas:

- Pool resiliency is enhanced by mirroring the storage pool (two failover groups)
- Pool shrink is enhanced by allowing the dynamic contraction of the storage pool by removing a physical volume
- Dynamic disk growth within the storage pool
- Scaling improvements with more client VMs supported and larger physical volumes in the pool
- New **lu** and **pv** commands
- New **failgrp** command
- Cluster-wide operations performed concurrently

As a reference, the following list includes VIOS Version 2.2.2 enhancements:

- Rolling updates that allow software updates to be applied sequentially to the VIOSs without needing to stop the whole cluster
- Repository resiliency that allows the cluster to remain operational even though the repository disk has failed. While the repository disk is in the failed state, configuration changes to the cluster cannot be made, but the cluster remains operational for client VMs. Also, the repository disk can be replaced.
- Virtual Network VLAN tagging support in the cluster

4.2.1 Shared Storage Pool

A Shared Storage Pool (SSP) is a server-based storage virtualization clustered solution and is an extension of existing storage virtualization options on the VIOS. An SSP can simplify the aggregation of large numbers of disks. It also allows better utilization of the available storage by using thin provisioning. The thinly provisioned device is not fully backed by physical storage if the data block is not actually in use. Thick (or full) provisioning is also available. The SSP also provides a good storage setup for Live Partition Mobility.

Figure 4-4 on page 48 provides a high-level overview of SSP functionality.

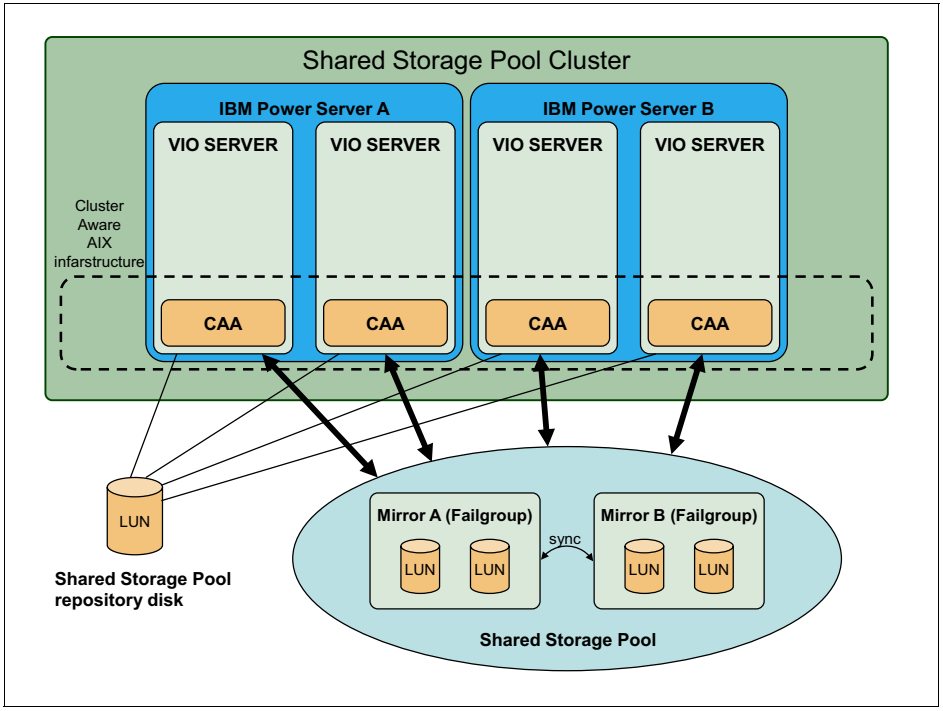


Figure 4-4 SSP design overview

An SSP usually spans multiple VIOSs. The VIOSs constitute a cluster that is based on Cluster Aware AIX (CAA) technology in the background. A cluster manages a single SSP. After the physical volumes are allocated to the SSP environment, the physical volume management tasks, such as capacity management, are performed by the cluster. Physical storage that becomes part of the SSP in these VIOSs is no longer managed locally.

SSPs provide the following benefits:

- Simplify the aggregation of large numbers of disks across multiple VIOSs
- Improve the utilization of the available storage
- Simplify administration tasks
- Provides global view of the storage across all VIOSs in the cluster
- Provides simple and error proof environment for Live Partition Mobility

SSP storage access

When using SSPs, the VIOS provides storage through logical units (LUs) that are assigned to client partitions. A logical unit (LU) is a file backed storage device that resides in the cluster file system in the SSP. It is mapped over a vSCSI adapter pair to a client virtual machine and it appears as a virtual SCSI disk in the client VM.

Dual VIOS configuration for high resiliency is supported. An LU can be propagated to a VM by multiple VIOSs and the client VM will access it by native MPIO driver.

Figure 4-5 shows how data is accessed from client VM through all layers to the physical storage.

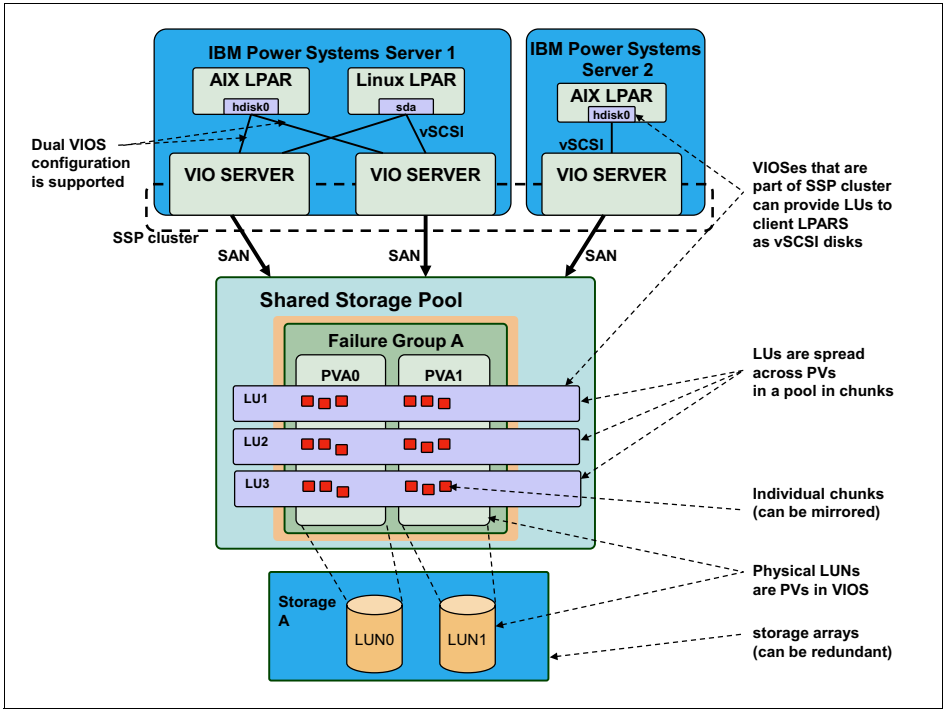


Figure 4-5 SSP data flow

The virtual SCSI disk devices exported from the SSP support SCSI persistent reservations. These SCSI persistent reservations persist across (hard) resets. The persistent reservations supported by a virtual SCSI disk from the SSP support all the required features for the SCSI-3 Persistent Reserves standard.

Note: Because IBM AIX operating system does not support full SCSI-3 persistent reserve capabilities, SSP implements additional options. *PR_exclusive* and *PR_shared* options are added to the reserve policy for virtual SCSI disk devices. The PR_exclusive is a persistent reserve for exclusive host access, and the PR_shared is a persistent reserve for shared host access. SSP technology provides functionality for these options.

SSP failure groups

Mirroring an SSP is a new feature in VIOS Version 2.2.3. Mirroring an SSP is an optional step that increases resiliency by adding redundancy.

Inside the storage pool, there might be two sets of shared LUNs (physical volumes (PVs)). These two named sets of LUNs are referred to as *failure groups* or *mirrors*. The preferred practice is to define those two failure groups on different physical storage arrays for best availability.

The whole pool is either a single copy pool (one failure group) or double copy (two failure groups). If two failure groups are defined, the whole pool is mirrored, not just individual logical units (LUs) of PVs. Data space that belongs to an LU is divided into 64 MB chunks each and they are placed into individual physical volumes (LUNs) in the pool. The exact data placement is decided in the background; therefore, it is not exact physical one-to-one mirroring (like RAID1, for example).

By default, a single copy pool is created by the **`cluster -create`** command with first failure group named `Default`. It is possible to rename the first failure group to an arbitrary name and add a second failure group.

Consider the following characteristics of mirrored SSP:

- It doubles the disk space requirement, which is typical for Disaster Recovery (DR) solutions.
- It is completely transparent for client VMs; therefore, there is no action needed on the client operating system. VIOS does all the work to access storage and to keep mirrors in a synchronized state. It duplicates writes to both mirrors and does re-mirroring if one of the mirrors becomes out-of-sync.
- VIOS performs recovery and re-mirroring automatically in the background without affecting the client VMs.

The following preferred practices relate to mirrored storage pools:

- Failure groups need to be kept the same size. If there are two failure groups in an SSP and their capacity is not the same, the total size of the SSP available for allocation of LUs is the sum of capacity of LUNs that are in the smaller failure group. The rest of the capacity in the larger failure group is not used.
- When creating a large mirrored pool with two failure groups, the preferred practice is to create a pool of one disk and add the second failure group to mirror the pool. After that, you can add physical volumes to both failure groups to increase the capacity of the pool.
- If a disk or a storage controller in a single failure group fails, the mirrored storage pool is running in a degraded state. Corrective actions to resolve the issue on the storage controller need to be taken.
- For the best mirroring performance, upgrade system firmware to the latest release.

Figure 4-6 shows the placement and flow of data when you use SSP mirroring.

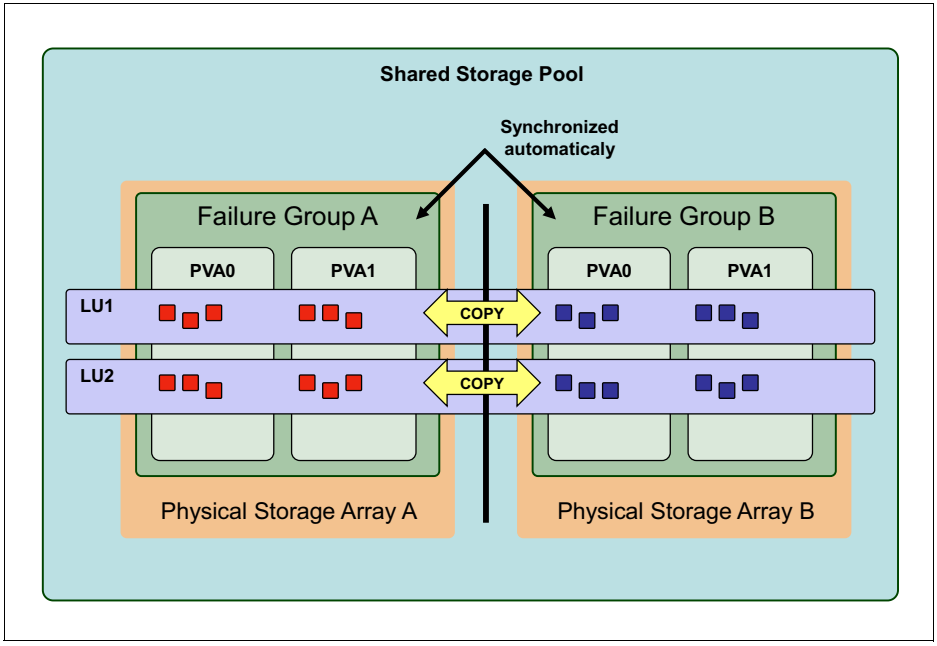


Figure 4-6 SSP mirroring

Clustering model

The underlying technology for clustering in an SSP is provided by the Cluster Aware AIX (CAA) component. Each VIOS that is part of a cluster represents a cluster node.

The storage in an SSP is managed by the cluster and a distributed data object repository with a global namespace. The distributed data object repository uses a cluster file system that has been enhanced specifically for the purpose of storage virtualization using the VIOS. The distributed object repository is the foundation for advanced storage virtualization features, such as shared access, thin provisioning, and mirroring.

The VIOS clustering model is based on Cluster Aware AIX (CAA) and Reliable Scalable Cluster Technology (RSCT). CAA is a toolkit for creating clusters. A reliable network connection is needed between all the VIOSs that are in the cluster. On the VIOS, the **poold** daemon handles group services. The **vio_daemon** is responsible for monitoring the health of the cluster nodes and the pool, as well as the pool capacity.

CAA provides a set of tools and APIs to enable clustering on the AIX operating system (which is the base of the VIOS appliance). CAA does not provide the application monitoring and resource failover capabilities that IBM PowerHA® System Mirror provides. Other software products can use the APIs and command-line interfaces (CLIs) that CAA provides to cluster their applications and services.

The following products use the CAA technology:

- RSCT (3.1 and later)
- PowerHA (7.1 and later)
- VIOS SSP (since 2.2.0.11, Fix Pack (FP) 24, Service Package (SP) 01)

Each cluster based on CAA requires at least one physical volume for the metadata repository. All cluster nodes in a cluster must see all the shared disks - both repository disk and storage disks. Therefore, the disks need to be zoned and correctly masked on the storage array to all the cluster nodes that are part of the SSP. All nodes can read and write to the SSP. The cluster uses a distributed lock manager to manage access to the storage.

Nodes that belong to a CAA cluster use the common AIX HA File System (AHAFS) for event notification. AHAFS is a pseudo file system used for synchronized information exchange; it is implemented in the AIX kernel extension.

CAA features

CAA offers these features:

- Central repository (repository disk - `caavg_private` volume group)
- Quorumless (CAA does not require a quorum to be up and operational.)
- Monitoring capabilities for custom actions
- Fencing for these areas:
  - Network
  - Storage
  - Applications
- Deadman switch (DMS). A *deadman switch* is an action that occurs when CAA detects that a node is isolated in a multinode environment. No network or disk communication occurs between nodes. Implement the DMS to protect the data on the external disks.
- CAA DMS tunable (**`deadman_mode`**) allows two actions:
  - Assert (crash) the system - default setting
  - Generate an AHAFS event

AIX HA File System

AIX HA File System (AHAFS) is implemented as a kernel extension. AHAFS is mounted on `/aha`. It can monitor predefined and user-defined system events. AHAFS automatically notifies registered users or processes about the occurrences of the following types of events:

- Modification of content of a file
- Usage of a file system that exceeds a user-defined threshold
- Death of a process
- Change in the value of a kernel tunable parameter

AHAFS offers these key features:

- No new API for monitoring events is needed. The monitoring applications just need to use the existing file system interfaces (for example, open(), write(), select(), read(), and close()).
- The same event can be monitored by many users or processes, each with a different threshold.
- Different levels of information can be extracted by the different users or processes upon the occurrence of an event.
- Any component or subcomponent in the kernel space, including kernel extensions and device drivers, can register its own event producers to AHAFS to enable the monitoring of its events.

CAA reports the following events via AHAFS:

- Node/host: nodeList, nodeState, nodeContact, linkedCl, and nodeAddress (see `/aha/cluster`)
- Network: networkAdapterState (see `/aha/cluster`)
- Disk: diskState, clDiskList, clDiskState, repDiskState, and vgState (see `/aha/disk`)

4.2.2 Planning for SSPs

SSP clustering is continuously enhanced. In the latest version of VIOS 2.2.3, SSP supports larger disks in the pool and for more client VMs. Also, the underlying cluster infrastructure is changed. Cluster communication has been changed from multicast to unicast; therefore, cluster setup has been simplified and made less error-prone.

Table 4-1 lists the historical development of SSP cluster scalability.

Table 4-1 Cluster scaling in various versions of VIOS

SSP release	VIOS version	Maximum number of nodes in cluster
1	2.2.0.11, Fix Pack 24, Service Pack 1, and 2.2.1.0	1
2	2.2.1.3	4
3	2.2.2.0	16
4	2.2.3.0	16

Table 4-2 on page 55 lists the differences in various SSP parameters to previous versions of SSP.

Table 4-2 Capacity and scaling parameters in the latest release of SSP

Maximum value	VIO 2.2.2 (SSP Release 3)	VIO 2.2.3 (SSP Release 4)
Number of VIOS nodes in cluster	16	16
Number of physical disks in pool	1024	1024
Number of virtual disks (LUs) mappings in pool	8192	8192
Number of client LPARs per VIOS node	200	250
Capacity of physical disks in pool	4 TB	16 TB
Total storage capacity of storage pool	512 TB	512 TB
Maximum capacity of a virtual disk (LU) in pool	4 TB	4 TB
Number of repository disks	1	1

Table 4-3 on page 56 lists the requirements for installation and the use of the latest version of SSP functionality.

Table 4-3 Version 2.2.3 SSP requirements

Category	Required
Server hardware	IBM POWER6®, POWER7, or POWER7+ server or blade
Firmware	No requirement but the latest firmware (770 or 780) improves the performance of mirrored writes in an SSP pool
HMC	7.7.4 and later for a graphical interface to the VIOS
VIOS LPAR CPU	1 (Entitlement or Dedicated)
VIOS LPAR memory	4 GB
Adapter	At least one Fibre Channel adapter
Disks	Shared disks from SAN-attached storage: ► One for repository xx GB ► Other LUNs for SSP data (at least one)

For storage hardware that is supported in VIOS, see this website:

http://bit.ly/1oIPmLS

4.2.3 Installing SSPs

There are no specific steps to install the SSP feature because it is an internal feature of VIOS 2.2.3. Therefore, after the standard installation of VIOS 2.2.3, the SSP feature is immediately available.

4.2.4 Setting up SSPs

To set up a cluster, there must be at least one reliable network connection between all VIOSs that will become part of the cluster. Also, there must be one SAN LUN accessible from all VIOSs that will be used as the cluster repository.

If the Domain Name System (DNS) server is in place, the name resolution needs to use local resolution first. The order of name lookup can be configured in the `/etc/netsvc.conf` file (must be edited as root). The host table for local name resolution, `/etc/hosts`, needs to be correctly configured with the long names of all the VIOSs that will be part of the cluster. Both forward resolution and reverse resolution need to work correctly. It is advised to synchronize the clock among all VIOSs by Network Time Protocol (NTP). The Shared Ethernet Adapter (if used) must be in the default threaded mode.

If all previous setup steps are completed and all the planning requirements that are described in 4.2.2, “Planning for SSPs” on page 54 are met, it is possible to create a cluster using the **`cluster`** command. The initial cluster setup can take time. There is not much feedback on the window while the cluster is being created.

Example 4-1 SSP cluster -create command

```
cluster -create -clustername SSP -repopvs ssprepohdisk0 -spname SSPpool
-sppvs sspmirrahdisk0 -hostname vioa1.pwrvc.ibm.com
Cluster SSP has been created successfully.
```

Note: In the previous example and all the following examples in this chapter, we use custom logical names of physical disks. It is for the convenience of the administrator to have both consistent and meaningful logical device names across the entire cluster. The renaming is done by running the **`rendev`** command (as root). This step is optional and not necessary in cluster configuration:

```
vioa1.pwrvc.ibm.com:/# rendev -l hdisk6 -n ssprepohdisk0
vioa1.pwrvc.ibm.com:/# rendev -l hdisk7 -n sspmirrahdisk0
```

We can check whether the cluster is defined successfully by using the **`cluster -status`** and **`lscluster -d`** commands as shown in Example 4-2.

Example 4-2 Cluster status listing

```
$ cluster -status -clustername SSP
Cluster Name         State
SSP                  OK

    Node Name        MTM           Partition Num  State  Pool State
    vioa1            8205-E6C02O6A22ER         1  OK     OK
$ lscluster -d
Storage Interface Query

Cluster Name: SSP
```

```
Cluster UUID: a8035a02-3c0e-11e3-9cb8-e41f13fdcf7c
Number of nodes reporting = 1
Number of nodes expected = 1
Node vioa1.pwrvc.ibm.com
Node UUID = a806fb6c-3c0e-11e3-9cb8-e41f13fdcf7c
Number of disks discovered = 2
        sspmirrahdisk0:
              State : UP
               uDid :
332136005076801910 26C400000000000002404214503IBMfcp
               uUid : 8290a634-5275-645d-9203-478a0e090ee1
          Site uUid : a8028ac8-3c0e-11e3-9cb8-e41f13fdcf7c
               Type : CLUSDISK
        ssprepohdisk0:
              State : UP
               uDid :
332136005076801910 26C400000000000002304214503IBMfcp
               uUid : 7fbcc0ec-e0ec-9127-9d51-96384a17c9d7
          Site uUid : a8028ac8-3c0e-11e3-9cb8-e41f13fdcf7c
               Type : REPDISK
```

If everything works well, all defined disks are visible, the cluster is in an `OK` state, all disks are `UP`, we can continue adding nodes (Example 4-3).

Example 4-3 SSP cluster addnode

```
vioa1:/home/padmin [padmin]$ cluster -addnode -clustername SSP
-hostname vioa2.pwrvc.ibm.com
Partition vioa2.pwrvc.ibm.com has been added to the SSP cluster.
```

And again, check the status (Example 4-4).

Example 4-4 SSP cluster status

```
$ cluster -status -clustername SSP
Cluster Name         State
SSP                  OK

    Node Name        MTM           Partition Num  State  Pool State
    vioa1            8205-E6C0206A22ER         1  OK     OK
    vioa2            8205-E6C0206A22ER         2  OK     OK
```

4.2.5 SSP management

Commands that were added or enhanced in the last version of SSP are described.

New lu command

The new **lu** command is introduced to simplify the management of the logical units within an SSP. By using the **lu** command, various operations, such as create, map, unmap, remove, and list, can be performed on logical units in an SSP.

The following list shows various flags of the **lu** command and examples of its usage:

- **-create** creates a new logical unit. By default, a thin-provisioned logical unit is created. Use the **-thick** option to create a thick-provisioned logical unit. Use the **-map** flag to map an existing logical unit to a virtual SCSI adapter. An example of the usage follows:

  ```
  $ lu -create -clustername SSP -sp SSPpool -lu vmaix10_hd0 -size 20G
  Lu Name:vmaix10_hd0
  Lu Udid:461b48367543c261817e3c2cfc326d12
  ```

- **-list** displays information about the logical units in the SSP. Use the **-verbose** option to display the detailed information about logical units. An example of the usage follows:

  ```
  $ lu -list
  POOL_NAME: SSPpool
  TIER_NAME: SYSTEM
  LU_NAME                 SIZE(MB)    UNUSED(MB)  UDID
  vmaix10_hd0             20480       20481
  461b48367543c261817e3c2cfc326d12
  ```

- **-map** maps an existing LU to the virtual target adapter. An example of the usage follows:

  ```
  $ lu -map -clustername SSP -sp SSPpool -lu vmaix10_hd0 -vadapter
  vhost8 -vtd vtd_vmaix10_hd0
  Assigning logical unit 'vmaix10_hd0' as a backing device.
  VTD:vtd_vmaix10_hd0
  ```

- **-unmap** unmaps an existing LU but does not delete it. An example of the usage follows:

  ```
  $ lu -unmap -clustername SSP -sp SSPpool -lu vmaix10_hd0
  vtd_vmaix10_hd0 deleted
  ```

- **-remove** removes the logical units from the SSP. To remove all the logical units, the optional flag **-all** can be used. An example of the usage follows:

```
$ lu -remove -clustername SSP -sp SSPpool -lu vmaix10_hd0
Logical unit vmaix10_hd0 with udid
"461b48367543c261817e3c2cfc326d12" is removed.
```

Important warning: Using the option **-all** will immediately delete all LUs in the pool even if they are mapped to a VM.

New pv command

The new **pv** command is introduced to manage the physical volumes (shared SAN LUNs) within an SSP. By using the **pv** command, various operations, such as add, add to a failure group, replace, remove, and list, can be performed on physical volumes in an SSP:

- **-list** lists physical volumes in an SSP and their Universal Disk Identification (UDIDs), for example:

```
$ pv -list
POOL_NAME: SSPpool
TIER_NAME: SYSTEM
FG_NAME: Default
PV_NAME          SIZE(MB)    STATE            UDID
sspmirrahdisk0   51200       ONLINE
332136005076801910 26C4000000000000~
```

- **-list -capable** lists the physical volumes that can be added to an SSP. The physical volumes that are accessible on all VIOSs across the entire cluster that are not part of the SSP will be listed. Also, UDIDs of those physical volumes will be listed:

```
$ pv -list -capable
PV_NAME          SIZE(MB)    UDID
sspmirrahdisk1   51200
332136005076801910 26C400000000000002504214503IBMfcp
sspmirrbhdisk0   51200
332136005076801910 26C400000000000002604214503IBMfcp
sspmirrbhdisk1   51200
332136005076801910 26C400000000000002704214503IBMfcp
```

- **-add** adds physical volumes to one or more failure groups in an SSP. When a disk is added to a storage pool, chunks that belong to already existing LUs in the pool are automatically redistributed in the background:

```
pv -add -fg MIRRA: sspmirrahdisk1 MIRRB: sspmirrbhdisk1
Given physical volume(s) have been added successfully.
```

- **-remove** removes physical volumes from an SSP. This reduces the capacity of the storage pool. This is a long running operation. The physical volume needs to be reused only after this operation completes successfully. An example follows:

```
$  pv -remove -clustername SSP -sp SSPpool -pv sspmirrahdisk1
Given physical volume(s) have been removed successfully.
```

- **-replace** replaces physical volumes in an SSP:

```
$ pv -replace -oldpv sspmirrahdisk0 -newpv sspmirrahdisk1
Current request action progress: % 5
Current request action progress: % 100
Given physical volume(s) have been replaced successfully.
```

New failgrp command

The **failgrp** command is used to manage the failure groups within an SSP. A *failure group* (a mirror) is a set of physical volumes that are treated as a single point of failure by the system. By using the **failgrp** command, various operations, such as create, remove, modify, and list, can be performed on the failure groups. When an SSP is created by using the **cluster -create** command, a single default failure group is created. A new failure group can later be created by using the **failgrp** command.

The following operations are performed by the **failgrp** command:

- **-create** creates a second failure group, which is a new failure group that mirrors the data in a default failure group. Ensure that the total capacity of the new failure group that is created is equal to or more than the capacity of the default failure group. An example follows:

```
$ failgrp -create -fg MIRRB: sspmirrbhdisk0
MIRRB FailureGroup has been created successfully.
```

- **-remove** removes a failure group from the SSP. Only one mirror copy of the data is removed. An example follows:

```
$ failgrp -remove -fg MIRRB
MIRRB FailureGroup has been removed successfully.
```

- **-list** displays information about the failure groups in an SSP:

```
$ failgrp -list
POOL_NAME: SSPpool
TIER_NAME: SYSTEM
FG_NAME                              FG_SIZE(MB) FG_STATE
MIRRA                                51136       ONLINE
MIRRB                                51136       ONLINE
$ failgrp -list -verbose
POOL_NAME:SSPpool
```

```
TIER_NAME:SYSTEM
FG_NAME:MIRRA
FG_SIZE(MB):51136
FG_STATE:ONLINE

POOL_NAME:SSPpool
TIER_NAME:SYSTEM
FG_NAME:MIRRB
FG_SIZE(MB):51136
FG_STATE:ONLINE
```

- **-modify** used together with the **-attr** flag modifies the specified attribute. The following example shows how to rename a default failure group to a new name:

```
$ failgrp -modify -fg Default -attr fg_name=MIRRA
Given attribute(s) modified successfully.
```

Note: Only two failure groups are currently supported in an SSP.

New chrepos command

This command replaces a disk, which is used as the repository disk by the SSP cluster, with another disk. It is not new in VIOS Version 2.2.3. It was introduced in Version 2.2.2, but it was enhanced to operate in a multinode cluster environment. Example 4-5 shows how to use it. It can also be used for recovery if a repository disk is lost. That procedure is shown next in 4.2.6, “SSP troubleshooting” on page 63.

Example 4-5 Using the chrepos command

```
vioa1: [padmin]$ lspv | grep caavg_private
ssprepohdisk0    00f7a22ec86a91ac caavg_private    active
vioa1: [padmin]$ chrepos -n SSP -r +ssprepohdisk1,-ssprepohdisk0
chrepos: Successfully modified repository disk or disks.
vioa1: [padmin]$ lspv | grep caavg_private
ssprepohdisk1    00f7a22e24d94c38 caavg_private    active
```

4.2.6 SSP troubleshooting

Troubleshooting the SSP and CCA infrastructure is described. It is not in the scope of this book to provide a full troubleshooting guide; therefore, we include only a few hints where to look when there are problems with SSP. The following commands might help you troubleshoot:

- **lscluster -i** (**lscluster -i | egrep 'Node|Interface'**)

  The command offers a quick overview of the cluster status and the status of all interfaces, both network and disk heartbeating. See Example 4-6.

  Example 4-6 SSP lscluster -i command

```
vioa1:/# lscluster -i | egrep 'Node|Interface'
Network/Storage Interface Query
Node vioa1.pwrvc.ibm.com
Node UUID = a806fb6c-3c0e-11e3-9cb8-e41f13fdcf7c
        Interface number 1, en6
                Interface state = UP
        Interface number 2, dpcom
                Interface state = UP RESTRICTED AIX_CONTROLLED
Node vioa2.pwrvc.ibm.com
Node UUID = 4e28a148-3c10-11e3-9177-e41f13fdcf7c
        Interface number 1, en6
                Interface state = UP
        Interface number 2, dpcom
                Interface state = UP RESTRICTED AIX_CONTROLLED
Node viob1.pwrvc.ibm.com
Node UUID = 2eddfec2-3c11-11e3-8be2-e41f13fdcf7c
        Interface number 1, en4
                Interface state = UP
        Interface number 2, dpcom
                Interface state = UP RESTRICTED AIX_CONTROLLED
Node viob2.pwrvc.ibm.com
Node UUID = adf248a8-3c11-11e3-8b0a-e41f13fdcf7c
        Interface number 1, en4
                Interface state = UP
        Interface number 2, dpcom
                Interface state = UP RESTRICTED AIX_CONTROLLED
```

- **lscluster -m**

  This command lists the cluster node configuration information together with a listing of contact IP addresses for individual nodes in the cluster. See Example 4-7 on page 64.

Example 4-7 SSP lscluster -m command

```
vioa1.pwrvc.ibm.com:/# lscluster -m
Calling node query for all nodes...
Node query number of nodes examined: 4

        Node name: vioa1.pwrvc.ibm.com
        Cluster shorthand id for node: 1
        UUID for node: a806fb6c-3c0e-11e3-9cb8-e41f13fdcf7c
        State of node: UP  NODE_LOCAL
        Smoothed rtt to node: 0
        Mean Deviation in network rtt to node: 0
        Number of clusters node is a member in: 1
        CLUSTER NAME       SHID         UUID
        SSP                0 a8035a02-3c0e-11e3-9cb8-e41f13fdcf7c
        SITE NAME          SHID         UUID
        LOCAL              1 a8028ac8-3c0e-11e3-9cb8-e41f13fdcf7c

        Points of contact for node: 0

----------------------------------------------------------------------

        Node name: vioa2.pwrvc.ibm.com
        Cluster shorthand id for node: 2
        UUID for node: 4e28a148-3c10-11e3-9177-e41f13fdcf7c
        State of node: UP
        Smoothed rtt to node: 9
        Mean Deviation in network rtt to node: 5
        Number of clusters node is a member in: 1
        CLUSTER NAME       SHID         UUID
        SSP                0 a8035a02-3c0e-11e3-9cb8-e41f13fdcf7c
        SITE NAME          SHID         UUID
        LOCAL              1 a8028ac8-3c0e-11e3-9cb8-e41f13fdcf7c

        Points of contact for node: 1

----------------------------------------------------------------------
        Interface     State  Protocol    Status     SRC_IP->DST_IP

----------------------------------------------------------------------
tcpsock->02   UP     IPv4         none 172.16.21.110->172.16.21.111
----------------------------------------------------------------------

        Node name: viob1.pwrvc.ibm.com
        Cluster shorthand id for node: 3
```

```
        UUID for node: 2eddfec2-3c11-11e3-8be2-e41f13fdcf7c
        State of node: UP
        Smoothed rtt to node: 7
        Mean Deviation in network rtt to node: 3
        Number of clusters node is a member in: 1
        CLUSTER NAME       SHID         UUID
        SSP                0 a8035a02-3c0e-11e3-9cb8-e41f13fdcf7c
        SITE NAME          SHID         UUID
        LOCAL              1 a8028ac8-3c0e-11e3-9cb8-e41f13fdcf7c

        Points of contact for node: 1

-------------------------------------------------------------------
        Interface     State  Protocol    Status     SRC_IP->DST_IP
-------------------------------------------------------------------
tcpsock->03   UP     IPv4         none 172.16.21.110->172.16.21.112
-------------------------------------------------------------------

        Node name: viob2.pwrvc.ibm.com
        Cluster shorthand id for node: 4
        UUID for node: adf248a8-3c11-11e3-8b0a-e41f13fdcf7c
        State of node: UP
        Smoothed rtt to node: 7
        Mean Deviation in network rtt to node: 3
        Number of clusters node is a member in: 1
        CLUSTER NAME       SHID         UUID
        SSP                0 a8035a02-3c0e-11e3-9cb8-e41f13fdcf7c
        SITE NAME          SHID         UUID
        LOCAL              1 a8028ac8-3c0e-11e3-9cb8-e41f13fdcf7c

        Points of contact for node: 1

-------------------------------------------------------------------
        Interface     State  Protocol    Status     SRC_IP->DST_IP
-------------------------------------------------------------------
tcpsock->04   UP     IPv4         none 172.16.21.110->172.16.21.113
```

- **`lscluster -d`**

  This command shows list of disks currently configured in the cluster and their status. See Example 4-8 on page 66.

Example 4-8 SSP lscluster -d command

```
vioa1.pwrvc.ibm.com:/# lscluster -d
Storage Interface Query

Cluster Name: SSP
Cluster UUID: a8035a02-3c0e-11e3-9cb8-e41f13fdcf7c
Number of nodes reporting = 4
Number of nodes expected = 4

Node vioa1.pwrvc.ibm.com
Node UUID = a806fb6c-3c0e-11e3-9cb8-e41f13fdcf7c
Number of disks discovered = 2
        sspmirrahdisk1:
              State : UP
               uDid :
332136005076801910 26C400000000000002504214503IBMfcp
               uUid : b17cf1df-5ba1-38b6-9fbf-f7b1618a9010
          Site uUid : a8028ac8-3c0e-11e3-9cb8-e41f13fdcf7c
               Type : CLUSDISK
        ssprepohdisk0:
              State : UP
               uDid :
332136005076801910 26C400000000000002304214503IBMfcp
               uUid : 7fbcc0ec-e0ec-9127-9d51-96384a17c9d7
          Site uUid : a8028ac8-3c0e-11e3-9cb8-e41f13fdcf7c
               Type : REPDISK

Node viob2.pwrvc.ibm.com
Node UUID = adf248a8-3c11-11e3-8b0a-e41f13fdcf7c
Number of disks discovered = 2
        sspmirrahdisk1:
              State : UP
               uDid :
332136005076801910 26C400000000000002504214503IBMfcp
               uUid : b17cf1df-5ba1-38b6-9fbf-f7b1618a9010
          Site uUid : a8028ac8-3c0e-11e3-9cb8-e41f13fdcf7c
               Type : CLUSDISK
        ssprepohdisk0:
              State : UP
               uDid :
332136005076801910 26C400000000000002304214503IBMfcp
               uUid : 7fbcc0ec-e0ec-9127-9d51-96384a17c9d7
          Site uUid : a8028ac8-3c0e-11e3-9cb8-e41f13fdcf7c
               Type : REPDISK
```

```
Node vioa2.pwrvc.ibm.com
Node UUID = 4e28a148-3c10-11e3-9177-e41f13fdcf7c
Number of disks discovered = 2
        sspmirrahdisk1:
              State : UP
               uDid :
332136005076801910260400000000000002504214503IBMfcp
               uUid : b17cf1df-5ba1-38b6-9fbf-f7b1618a9010
          Site uUid : a8028ac8-3c0e-11e3-9cb8-e41f13fdcf7c
               Type : CLUSDISK
        ssprepohdisk0:
              State : UP
               uDid :
332136005076801910260400000000000002304214503IBMfcp
               uUid : 7fbcc0ec-e0ec-9127-9d51-96384a17c9d7
          Site uUid : a8028ac8-3c0e-11e3-9cb8-e41f13fdcf7c
               Type : REPDISK

Node viob1.pwrvc.ibm.com
Node UUID = 2eddfec2-3c11-11e3-8be2-e41f13fdcf7c
Number of disks discovered = 2
        sspmirrahdisk1:
              State : UP
               uDid :
332136005076801910260400000000000002504214503IBMfcp
               uUid : b17cf1df-5ba1-38b6-9fbf-f7b1618a9010
          Site uUid : a8028ac8-3c0e-11e3-9cb8-e41f13fdcf7c
               Type : CLUSDISK
        ssprepohdisk0:
              State : UP
               uDid :
332136005076801910260400000000000002304214503IBMfcp
               uUid : 7fbcc0ec-e0ec-9127-9d51-96384a17c9d7
          Site uUid : a8028ac8-3c0e-11e3-9cb8-e41f13fdcf7c
               Type : REPDISK
```

- **lscluster -s**

  This command lists the cluster network statistics on the local node and errors in the network (if they occur). See Example 4-9 on page 68.

Example 4-9 SSP lscluster -s command

```
vioa1:/# lscluster -s
Cluster Network Statistics:

pkts seen: 5410487                    passed: 5410410
IP pkts: 5406267                      UDP pkts: 130752
gossip pkts sent: 3174                gossip pkts recv: 9537
cluster address pkts: 0               CP pkts: 23201
bad transmits: 11                     bad posts: 7
Bad transmit (overflow - disk ): 0
Bad transmit (overflow - tcpsock): 0
Bad transmit (host unreachable): 0
Bad transmit (net unreachable): 0
Bad transmit (network down): 3
Bad transmit (no connection): 0
short pkts: 0                         multicast pkts: 63
cluster wide errors: 0                bad pkts: 0
dup pkts: 94                          dropped pkts: 0
pkt fragments: 27                     fragments queued: 0
fragments freed: 0
pkts pulled: 0                        no memory: 0
rxmit requests recv: 4                requests found: 4
requests missed: 0                    ooo pkts: 0
requests reset sent: 0                reset recv: 0
remote tcpsock send: 23028            tcpsock recv: 46080
rxmit requests sent: 0
alive pkts sent: 0                    alive pkts recv: 0
ahafs pkts sent: 23                   ahafs pkts recv: 22
nodedown pkts sent: 0                 nodedown pkts recv: 0
socket pkts sent: 148                 socket pkts recv: 154
cwide pkts sent: 1649                 cwide pkts recv: 1819
socket pkts no space: 0               pkts recv notforhere: 40
Pseudo socket pkts sent: 0            Pseudo socket pkts recv: 0
Pseudo socket pkts dropped: 0
arp pkts sent: 11                     arp pkts recv: 5
stale pkts recv: 3                    other cluster pkts: 2
storage pkts sent: 1                  storage pkts recv: 1
disk pkts sent: 72                    disk pkts recv: 63
unicast pkts sent: 1735               unicast pkts recv: 428
out-of-range pkts recv: 0
IPv6 pkts sent: 0                     IPv6 pkts recv: 70
IPv6 frags sent: 0                    IPv6 frags recv: 0
Unhandled large pkts: 0
mrxmit overflow     : 0               urxmit overflow: 0
```

- **lsattr -El cluster0**

  This command lists local cluster IDs. See Example 4-10.

Example 4-10 SSp lsattr cluster command

```
vioa1:/# lsattr -El cluster0
clvdisk   7fbcc0ec-e0ec-9127-9d51-96384a17c9d7 Cluster repository
disk identifier True
node_uuid a806fb6c-3c0e-11e3-9cb8-e41f13fdcf7c OS image identifier
True
```

- **lssrc -ls cthags**

  This command shows the status of Reliable Scalable Cluster Technology (RSCT) cthags services. See Example 4-11.

Example 4-11 lssrc -ls cthags command

```
vioa1:/# lssrc -ls cthags
Subsystem         Group            PID          Status
 cthags           cthags           8257746      active
5 locally-connected clients.  Their PIDs:
8454168(IBM.ConfigRMd) 7340256(vio_daemon) 7864338(poold)
7012574(rmcd) 10682550(IBM.StorageRMd)
HA Group Services domain information:
Domain established by node 3
Number of groups known locally: 4
                   Number of   Number of local
Group name         providers   providers/subscribers
gsPool.SSP               4           1           2
rmc_peers                4           1           0
IBM.ConfigRM             4           1           0
IBM.StorageRM.v1         4           1           0

Critical clients will be terminated if unresponsive

Dead Man Switch Disabled
```

- **lssrc -ls vio_daemon**

  This command shows the status of the vio_daemon service. See Example 4-12 on page 70.

Example 4-12 SSP vio_daemon

```
vvioa1.pwrvc.ibm.com:/# lssrc -ls vio_daemon
Node ID:                        a806fb6c3c0e11e39cb8e41f13fdcf7c
Log File:                       /home/ios/logs/viod.log
VSP Socket:                     0
AF family:                      0
Port:
Addr:
VKE Kernel Socket:              4
VKE Daemon Socket:              5
Bound to :                      /home/ios/socks/vioke_unix
API Socket:                     7
Bound to :                      /home/ios/socks/api_eve_unix
Cluster Name:                   SSP
Cluster ID:                     a8035a023c0e11e39cb8e41f13fdcf7c
PNN NODE ID:                    00000000000000000000000000000000
DBN NODE ID:                    adf248a83c1111e38b0ae41f13fdcf7c
Pool Label:                     SSPpool
Pool VIO Name:                  D_E_F_A_U_L_T_061310
Pool ID:                        FFFFFFFFAC10156E0000000052681193
Pool State:                     UP
Pool Sync Status:               COMPLETED
Repository Cluster Mode:        EVENT
Repository Disk State:          UP
DBN Role:                       Other
PNN Role:                       Other
```

- **lssrc -ls IBM.StorageRM**

  This command shows the status of StorageRM (Resource Monitor) objects.

- **lssrc -ls IBM.ConfigRM**

  This command shows the status of ConfigRM (Resource Monitor) objects.

- **lscluster -c**

  This command lists the cluster configuration. See Example 4-13.

Example 4-13 SSP lscluster -c command

```
viob1.pwrvc.ibm.com:/# lscluster -c
Cluster Name: SSP
Cluster UUID: a8035a02-3c0e-11e3-9cb8-e41f13fdcf7c
Number of nodes in cluster = 4
        Cluster ID for node vioa1.pwrvc.ibm.com: 1
        Primary IP address for node vioa1.pwrvc.ibm.com:
172.16.21.110
```

```
        Cluster ID for node vioa2.pwrvc.ibm.com: 2
        Primary IP address for node vioa2.pwrvc.ibm.com:
172.16.21.111
        Cluster ID for node viob1.pwrvc.ibm.com: 3
        Primary IP address for node viob1.pwrvc.ibm.com:
172.16.21.112
        Cluster ID for node viob2.pwrvc.ibm.com: 4
        Primary IP address for node viob2.pwrvc.ibm.com:
172.16.21.113
Number of disks in cluster = 2
        Disk = sspmirrahdisk1 UUID =
b17cf1df-5ba1-38b6-9fbf-f7b1618a9010 cluster_major = 0 cluster_minor
= 3
        Disk = ssprepohdisk0 UUID =
7fbcc0ec-e0ec-9127-9d51-96384a17c9d7 cluster_major = 0 cluster_minor
= 1
Multicast for site LOCAL: IPv4 228.16.21.110 IPv6 ff05::e410:156e
Communication Mode: unicast
Local node maximum capabilities: HNAME_CHG, UNICAST, IPV6, SITE
Effective cluster-wide capabilities: HNAME_CHG, UNICAST, IPV6, SITE
```

- **cluster -status** command (also with **-verbose** option)
- This command shows the cluster status. See Example 4-14.

Example 4-14 SSP cluster status

```
vioa1 [padmin]$ cluster -status -clustername SSP
Cluster Name         State
SSP                  OK

    Node Name        MTM           Partition Num  State  Pool State
    vioa1            8205-E6C0206A22ER         1  OK     OK
    vioa2            8205-E6C0206A22ER         2  OK     OK
    viob1            8233-E8B02061AA6P         1  OK     OK
    viob2            8233-E8B02061AA6P         2  OK     OK
vioa1 [padmin]$ cluster -status -clustername SSP -verbose
Cluster Name:             SSP
Cluster Id:               a8035a023c0e11e39cb8e41f13fdcf7c
Cluster State:            OK
Repository Mode:          EVENT
Number of Nodes:          4
Nodes OK:                 4
Nodes DOWN:               0

    Pool Name:            SSPpool
```

```
Pool Id:              FFFFFFFFAC10156E0000000052681193
Pool Mirror State:    NOT_MIRRORED

Node Name:            vioa1.pwrvc.ibm.com
Node Id:              a806fb6c3c0e11e39cb8e41f13fdcf7c
Node MTM:             8205-E6C0206A22ER
Node Partition Num:   1
Node State:           OK
Node Repos State:     OK
Node Upgrade Status:  2.2.3.0 ON_LEVEL
Node Roles:
    Pool Name:        SSPpool
    Pool Id:          FFFFFFFFAC10156E0000000052681193
    Pool State:       OK

Node Name:            vioa2.pwrvc.ibm.com
Node Id:              4e28a1483c1011e39177e41f13fdcf7c
Node MTM:             8205-E6C0206A22ER
Node Partition Num:   2
Node State:           OK
Node Repos State:     OK
Node Upgrade Status:  2.2.3.0 ON_LEVEL
Node Roles:
    Pool Name:        SSPpool
    Pool Id:          FFFFFFFFAC10156E0000000052681193
    Pool State:       OK

Node Name:            viob1.pwrvc.ibm.com
Node Id:              2eddfec23c1111e38be2e41f13fdcf7c
Node MTM:             8233-E8B02061AA6P
Node Partition Num:   1
Node State:           OK
Node Repos State:     OK
Node Upgrade Status:  2.2.3.0 ON_LEVEL
Node Roles:
    Pool Name:        SSPpool
    Pool Id:          FFFFFFFFAC10156E0000000052681193
    Pool State:       OK

Node Name:            viob2.pwrvc.ibm.com
Node Id:              adf248a83c1111e38b0ae41f13fdcf7c
Node MTM:             8233-E8B02061AA6P
Node Partition Num:   2
Node State:           OK
Node Repos State:     OK
```

```
    Node Upgrade Status:  2.2.3.0 ON_LEVEL
    Node Roles:           DBN
        Pool Name:        SSPpool
        Pool Id:          FFFFFFFFAC10156E0000000052681193
        Pool State:       OK
```

- **chrepos**

  This command replaces a disk, which is used as the repository disk by the SSP cluster, with another disk. Example 4-15 shows how to use this command to recover if you have a lost repository disk.

Example 4-15 Using chrepos to replace a failed repository disk

```
vioa1:/home/padmin [padmin]$ lspv | grep caavg_private
ssprepohdisk1    00f7a22e24d94c38                    caavg_private
active
```

Now, we unmapped ssprepohdisk1 (**unmap all hosts** operation on SAN Volume Controller)

```
vioa1.pwrvc.ibm.com:/# lsvg -p caavg_private
0516-062 : Unable to read or write logical volume manager
        record. PV may be permanently corrupted. Run diagnostics

vioa1: [padmin]$ chrepos -n SSP -r +ssprepohdisk0,-ssprepohdisk1
ERROR: return = -1 ssprepohdisk1 is not a valid repository device.
ERROR: return = -1, Could not read from cluster repository device
/dev/rssprepohdisk1: The specified device does not exist.
ERROR: Not a valid cluster header.
ERROR: return = -1 ssprepohdisk1 is not a valid repository device.
ERROR: return = -1 ssprepohdisk1 is not a valid repository device.
WARNING: importvg on ssprepohdisk1 failed.
WARNING: reducevg on ssprepohdisk1 failed.
ERROR: return = -1, Could not open cluster repository device
/dev/rssprepohdisk1: There is an input or output error.
WARNING: Failed to read repository data.
ERROR: return = -1, Could not open cluster repository device
/dev/rssprepohdisk1: There is an input or output error.
WARNING: Failed to write repository data.
WARNING: Unable to destroy repository disk ssprepohdisk1. Manual
intervention is required to clear the disk of cluster identifiers.
chrepos: Successfully modified repository disk or disks.
```

Replace operation succeeded. All is OK in the cluster now.

```
vioa1: [padmin]$ lspv | grep caavg_private
```

```
ssprepohdisk0    00f7a22ec86a91ac                    caavg_private
active
vioa1:/home/padmin [padmin]$ lscluster -d
Storage Interface Query

Cluster Name: SSP
Cluster UUID: a8035a02-3c0e-11e3-9cb8-e41f13fdcf7c
Number of nodes reporting = 4
Number of nodes expected = 4

Node vioa1.pwrvc.ibm.com
Node UUID = a806fb6c-3c0e-11e3-9cb8-e41f13fdcf7c
Number of disks discovered = 2
        sspmirrahdisk1:
              State : UP
               uDid :
33213600507680191026C4000000000000002504214503IBMfcp
               uUid : b17cf1df-5ba1-38b6-9fbf-f7b1618a9010
          Site uUid : a8028ac8-3c0e-11e3-9cb8-e41f13fdcf7c
               Type : CLUSDISK
        ssprepohdisk0:
              State : UP
               uDid :
33213600507680191026C4000000000000002304214503IBMfcp
               uUid : 7fbcc0ec-e0ec-9127-9d51-96384a17c9d7
          Site uUid : a8028ac8-3c0e-11e3-9cb8-e41f13fdcf7c
               Type : REPDISK

Node viob2.pwrvc.ibm.com
Node UUID = adf248a8-3c11-11e3-8b0a-e41f13fdcf7c
Number of disks discovered = 2
        sspmirrahdisk1:
              State : UP
               uDid :
33213600507680191026C4000000000000002504214503IBMfcp
               uUid : b17cf1df-5ba1-38b6-9fbf-f7b1618a9010
          Site uUid : a8028ac8-3c0e-11e3-9cb8-e41f13fdcf7c
               Type : CLUSDISK
        ssprepohdisk0:
              State : UP
               uDid :
33213600507680191026C4000000000000002304214503IBMfcp
               uUid : 7fbcc0ec-e0ec-9127-9d51-96384a17c9d7
          Site uUid : a8028ac8-3c0e-11e3-9cb8-e41f13fdcf7c
               Type : REPDISK
```

```
Node vioa2.pwrvc.ibm.com
Node UUID = 4e28a148-3c10-11e3-9177-e41f13fdcf7c
Number of disks discovered = 2
        sspmirrahdisk1:
              State : UP
               uDid :
332136005076801910260400000000000002504214503IBMfcp
               uUid : b17cf1df-5ba1-38b6-9fbf-f7b1618a9010
          Site uUid : a8028ac8-3c0e-11e3-9cb8-e41f13fdcf7c
               Type : CLUSDISK
        ssprepohdisk0:
              State : UP
               uDid :
332136005076801910260400000000000002304214503IBMfcp
               uUid : 7fbcc0ec-e0ec-9127-9d51-96384a17c9d7
          Site uUid : a8028ac8-3c0e-11e3-9cb8-e41f13fdcf7c
               Type : REPDISK

Node viob1.pwrvc.ibm.com
Node UUID = 2eddfec2-3c11-11e3-8be2-e41f13fdcf7c
Number of disks discovered = 2
        sspmirrahdisk1:
              State : UP
               uDid :
332136005076801910260400000000000002504214503IBMfcp
               uUid : b17cf1df-5ba1-38b6-9fbf-f7b1618a9010
          Site uUid : a8028ac8-3c0e-11e3-9cb8-e41f13fdcf7c
               Type : CLUSDISK
        ssprepohdisk0:
              State : UP
               uDid :
332136005076801910260400000000000002304214503IBMfcp
               uUid : 7fbcc0ec-e0ec-9127-9d51-96384a17c9d7
          Site uUid : a8028ac8-3c0e-11e3-9cb8-e41f13fdcf7c
               Type : REPDISK
```

- **snap caa**

  Use this command to collect all information about the underlying CAA cluster component when sending information to IBM support.

- log **/var/adm/ras/syslog.caa**

  The CAA subsystem logs its events to this file.

- **/usr/lib/cluster/clras lsrepos**

  Use this command to list valid cluster repository disks. See Example 4-16.

Example 4-16 SSP clras

```
vioa1:/# /usr/lib/cluster/clras lsrepos
ssprepohdisk0 has a cluster repository signature.
Cycled 10 disks.
Found 1 cluster repository disk.
```

- **/usr/lib/cluster/clras sfwinfo -d hdiskx** displays the Universally Unique Identifiers (UUIDs) for disk. See Example 4-17.

Example 4-17 SSP clras sfwinfo

```
vioa1:/# /usr/lib/cluster/clras sfwinfo -d sspmirrahdisk0
sspmirrahdisk0  8290a634-5275-645d-9203-478a0e090ee1
```

- **/usr/lib/cluster/clras dumprepos** shows the content of the repository disk. See Example 4-18.

Example 4-18 SSP dumprepos

```
vioa1:/# /usr/lib/cluster/clras dumprepos
HEADER
        Cluster ID:     0xa9c2d4c2
        Name:           SSP
        UUID:           a8035a02-3c0e-11e3-9cb8-e41f13fdcf7c
        Checksum:       0xa6c0

SHARED DISKS
       Name                         Uuid
Udid
        sspmirrahdisk1
b17cf1df-5ba1-38b6-9fbf-f7b1618a9010
33213600507680191026C4000000000000002504214503IBMfcp

NODES
       Name                         Uuid
N_gw    Site_uuid
        vioa1.pwrvc.ibm.com
a806fb6c-3c0e-11e3-9cb8-e41f13fdcf7c    1
a8028ac8-3c0e-11e3-9cb8-e41f13fdcf7c
        gw_flag : 0
```

```
        gw_ip
        172.16.21.110
        vioa2.pwrvc.ibm.com
4e28a148-3c10-11e3-9177-e41f13fdcf7c    1
a8028ac8-3c0e-11e3-9cb8-e41f13fdcf7c
        gw_flag : 0
        gw_ip
        172.16.21.111
        viob1.pwrvc.ibm.com
2eddfec2-3c11-11e3-8be2-e41f13fdcf7c    1
a8028ac8-3c0e-11e3-9cb8-e41f13fdcf7c
        gw_flag : 0
        gw_ip
        172.16.21.112
        viob2.pwrvc.ibm.com
adf248a8-3c11-11e3-8b0a-e41f13fdcf7c    1
a8028ac8-3c0e-11e3-9cb8-e41f13fdcf7c
        gw_flag : 0
        gw_ip
        172.16.21.113

SITES
      Name                      Shid   Uuid
Prio
        LOCAL                           1
a8028ac8-3c0e-11e3-9cb8-e41f13fdcf7c    1


REPOS DISKS
       Name                         Uuid
Udid
        ssprepohdisk0
7fbcc0ec-e0ec-9127-9d51-96384a17c9d7
33213600507680191026C4000000000000023042145O3IBMfcp

MCAST ADDRS
        IPv4                    IPv6                    Uuid
        228.16.21.110           ff05::e410:156e
aa5566ec-3c0e-11e3-9cb8-e41f13fdcf7c
```

5

Virtual I/O Server Performance Advisor

This chapter describes how you can use the Virtual I/O Server (VIOS) Performance Advisor monitoring tool.

The VIOS Performance Advisor tool provides advisory reports based on key performance metrics for various partition resources collected from the VIOS environment.

This chapter includes the following sections:

- VIOS Performance Advisor concepts
- Using the VIOS Performance Advisor tool
- VIOS Performance Advisor reports

5.1 VIOS Performance Advisor concepts

The goal of the VIOS Performance Advisor is to have an expert system view of the performance metrics already available to you. The VIOS Performance Advisor helps you make assessments and recommendations based on the expertise and experience available within the IBM Systems Performance Group.

The VIOS Performance Advisor tool provides advisory reports that are based on key performance metrics of various partition resources collected from the VIOS environment. Use this tool to provide health reports that have proposals for making configurational changes to the VIOS environment and to identify areas for further investigation.

VIOS Version 2.2, Fix Pack (FP) 24, Service Package (SP) 1 includes the following enhancements for the VIOS Performance Advisor tool. However, the development of new functions for virtualization is an ongoing process. Therefore, it is best to visit the following website, where you can find more information about the new and existing features:

http://bit.ly/1nctYzk

The primary focus of the VIOS Performance Advisor is to cover the following VIOS technologies:

SEA	Shared Ethernet Adapter
NPIV	N_Port ID Virtualization
SSP	Shared Storage Pool

The VIOS Performance Advisor has been enhanced to provide support for NPIV and Fibre Channel, Virtual Networking, Shared Ethernet Adapter, and Shared Storage Pool configurations.

5.2 Using the VIOS Performance Advisor tool

Starting with VIOS Version 2.2.2.0 or later, you can use the VIOS Performance Advisor tool. By using the VIOS command-line interface (CLI), run the **part** command.

VIOS Performance Advisor can be downloaded for free from the VIOS Performance Advisor tool website, with VIOS Version 2.1.0.10, or later. By using the VIOS CLI, run the **vios_advisor** command.

For in-depth information, see the website:

http://www-01.ibm.com/support/docview.wss?uid=aixtools159f1226

You can start the VIOS Performance Advisor tool in the following modes with VIOS Version 2.2.2.0 or later:

- On-demand monitoring mode
- Postprocessing mode

5.2.1 On-demand monitoring mode

When you start the VIOS Performance Advisor tool in the on-demand monitoring mode, provide the duration for which the tool must monitor the system in minutes. The duration that you provide have to be between 10 - 60 minutes at the end of which the tool generates the reports. During this time, samples are collected at regular intervals of 15 seconds.

For example, to monitor the system for 30 minutes and generate a report, enter the following command:

```
vioa1:/home/padmin [padmin]$ part -i 10
part: Reports are successfully generated in vioa1_131031_11_34_12.tar
```

Note: If you use the **part** command, you must log in as the padmin role not the root user.

Reports for the on-demand monitoring mode are successfully generated in the `vioa1_131031_11_34_12.tar` file.

The output generated by the **part** command is saved in a `.tar` file, which is created in the current working directory. The naming convention for files in the on-demand monitoring mode is *`hostname_yymmdd_hhmmss.`*`tar`. In the postprocessing mode, the file name is that of the input file with the file name extension changed from a `.nmon` file to a `.tar` file.

The following example shows the tar file extracted from an output of the **part** command:

```
# ls
images                  vioa1_131031_1134.nmon   vios_advisorv2.xsl
popup.js                vios_advisor.xsl
style.css               vios_advisor_report.xml
```

The data is gathered in the `vioa1_131031_1134.nmon` file. If you want an `.xls` file for Excel, you can run the nmon_analyser.

For in-depth information about the nmon_analyser, see the website:

http://bit.ly/1o2Yhbc

5.2.2 Postprocessing mode

When you start the VIOS Performance Advisor tool in postprocessing mode, you must provide an input file. The tool tries to extract as much data as possible from the file that you provide. Then, the tool generates reports. If the input file does not have the required data for the tool to generate reports, an `Insufficient Data` message is added to the relevant fields. For example, to generate a report based on the data available in the `vioa1_131031_1134.nmon` file, enter the following command as shown in Example 5-1.

Example 5-1 The part command based on the .nmon file

```
vioa1:/home/padmin [padmin]$ part -f vioa1_131031_1134.nmon
part: Reports are successfully generated in vioa1_131031_1134.tar
```

Reports for the postprocessing mode are successfully generated in the `vioa1_131031_1134.tar` file. Extract the `vioa1_131031_1134.tar` file, and examine the `vios_advisor_report.xml` file in the extracted folder.

5.3 VIOS Performance Advisor reports

The VIOS Performance Advisor tool provides advisory reports that relate to the performance of various subsystems in the VIOS environment.

5.3.1 Transferring the .xml file to a browser-capable PC

The output generated by the **part** command is saved in a `.tar` file that is created in the current working directory.

The `vios_advisor.xml` report is part of the `output .tar` file with the other supporting files. To view the generated report, complete the following example.

```
vioa1:/home/padmin [padmin]$ ls -al vioa1_131031_11_34_12.tar
-rw-r--r-- 1 padmin staff 337920 Oct 31 11:44 vioa1_131031_11_34_12.tar
vioa1:/home/padmin [padmin]$ oem_setup_env
# cd /home/padmin
# tar -xf vioa1_131031_11_34_12.tar
# ls
images                  vioa1_131031_1134.nmon   vios_advisorv2.xsl
popup.js                vios_advisor.xsl
style.css               vios_advisor_report.xml
```

Then, download the `vios_advisor.xml`, and open the `vios_advisor.xml` file using a browser.

5.3.2 Advisory reports

The following types of advisory reports are generated by the VIOS Performance Advisor tool:

- System configuration advisory report
- CPU (central processing unit) advisory report
- Shared Processing Pool advisory report
- Memory advisory report
- Disk Drives advisory report
- Disk adapter advisory report
- I/O activities (disk and network) advisory report
- Shared Storage Pool advisory report
- Shared Ethernet Adapter advisory report

The Advisory report output is similar to Figure 5-1.

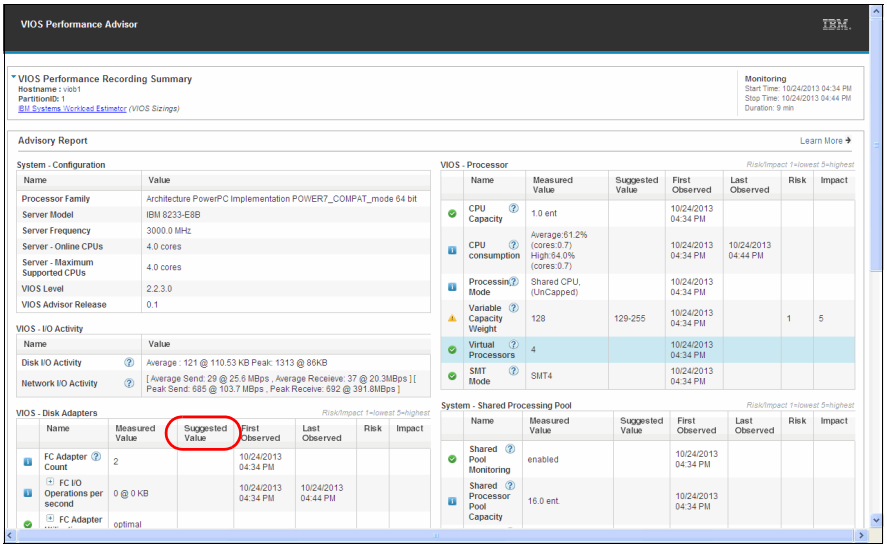


Figure 5-1 Advisory report information

Note: The Suggested value column (highlighted in Figure 5-1) shows changes that are advised to decrease performance risks and impacts.

Figure 5-2 shows the icon definitions.

Icons	Definitions
	Information related to configuration parameters
	Values acceptable in most cases
	Possible performance problem
	Severe performance problem
	Investigation required

Figure 5-2 VIOS Performance Advisor icons

5.3.3 The system configuration advisory report

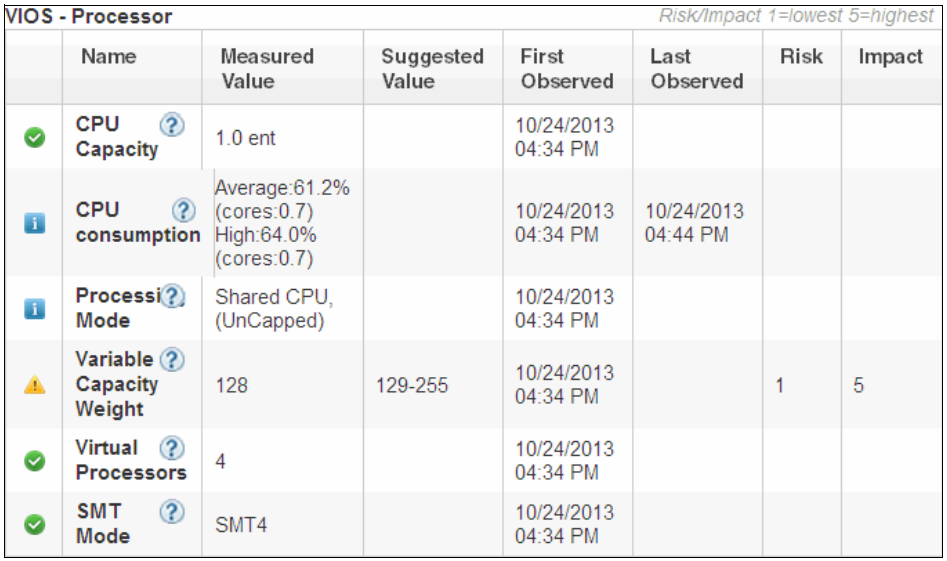
The System - Configuration advisory report consists of the information that relates to the VIOS configuration, such as processor family, server model, number of cores, frequency at which the cores are running, and the VIOS version. The output is similar to Figure 5-3.

System - Configuration	
Name	Value
Processor Family	Architecture PowerPC Implementation POWER7_COMPAT_mode 64 bit
Server Model	IBM 8233-E8B
Server Frequency	3000.0 MHz
Server - Online CPUs	4.0 cores
Server - Maximum Supported CPUs	4.0 cores
VIOS Level	2.2.3.0
VIOS Advisor Release	0.1

Figure 5-3 The system configuration advisory report

5.3.4 CPU (central processing unit) advisory report

The CPU or VIOS - Processor advisory report consists of the information that relates to the processor resources, such as the number of cores assigned to the VIOS and the processor consumption during the monitoring interval. The output is similar to Figure 5-4.



VIOS - Processor — Risk/Impact 1=lowest 5=highest

	Name	Measured Value	Suggested Value	First Observed	Last Observed	Risk	Impact
✓	CPU Capacity	1.0 ent		10/24/2013 04:34 PM			
i	CPU consumption	Average:61.2% (cores:0.7) High:64.0% (cores:0.7)		10/24/2013 04:34 PM	10/24/2013 04:44 PM		
i	Processi Mode	Shared CPU, (UnCapped)		10/24/2013 04:34 PM			
⚠	Variable Capacity Weight	128	129-255	10/24/2013 04:34 PM		1	5
✓	Virtual Processors	4		10/24/2013 04:34 PM			
✓	SMT Mode	SMT4		10/24/2013 04:34 PM			

Figure 5-4 CPU (central processing unit) advisory report

Note: In the VIOS - Processor table (Figure 5-4 on page 85) of the CPU (central processing unit) advisory report, the status of the Variable Capacity Weight is marked with a warning icon (exclamation point in a triangle). The preferred practice is for the VIOS to have an increased priority of 129 - 255 when in uncapped shared processor mode. For the definitions for the warning icons, see Figure 5-2 on page 84.

In Figure 5-5 on page 86, CPU capacity status indicates *investigation required*. For the VIOSs in our lab environment, the preferred practice capacity settings are used due to low performance requirements:

0.1 Processing units for desired entitled capacity
1 Desired virtual processor
255 Weight for uncapped processing mode

So, in this scenario, if a shortage of processing power occurs, all VIOSs still can use up to the suggested value of 0.7 Processing Units (PrU) and up to a whole physical processor. Therefore, the risk of this status is low and the impact is higher.

VIOS - Processor — *Risk/Impact 1=lowest 5=highest*

	Name	Measured Value	Suggested Value	First Observed	Last Observed	Risk	Impact
◆	CPU Capacity	0.1 ent	0.7 ent	10/28/2013 10:24 AM	10/28/2013 11:24 AM	1	4
i	CPU consumption	Average:612.3% (cores:0.7) High:670.0% (cores:0.7)		10/28/2013 10:24 AM	10/28/2013 11:24 AM		
i	Processi Mode	Shared CPU, (UnCapped)		10/28/2013 10:24 AM			
✓	Variable Capacity Weight	255		10/28/2013 10:24 AM			
✓	Virtual Processors	1		10/28/2013 10:24 AM			
✓	SMT Mode	SMT4		10/28/2013 10:24 AM			

Figure 5-5 CPU Capacity status in report

5.3.5 Shared Processing Pool advisory report

The System - Shared Processing Pool (SPP) advisory report consists of SPP resource-related information, such as shared processor pool capacity for virtual machines (VMs) running on the same processor pool. The output is similar to Figure 5-6. If there is free CPU capacity, the maximum capacity that an uncapped VM can use is the number of available virtual processors.

System - Shared Processing Pool — *Risk/Impact 1=lowest 5=highest*

	Name	Measured Value	Suggested Value	First Observed	Last Observed	Risk	Impact
✓	Shared Pool Monitoring	enabled		10/28/2013 10:24 AM			
i	Shared Processor Pool Capacity	16.0 ent.		10/28/2013 10:24 AM			
✓	Free CPU Capacity	average_free:15.1 ent. lowest_free:14.8 ent.					

Figure 5-6 Shared Processing Pool advisory report

In Figure 5-7, SPP monitoring is not enabled. The Shared Processor Pool monitoring feature must be enabled to view the statistics for an SPP.

System - Shared Processing Pool — Risk/Impact 1=lowest 5=highest

	Name	Measured Value	Suggested Value	First Observed	Last Observed	Risk	Impact
	Shared Pool Monitoring	disabled	enabled	10/31/2013 08:44 PM			
	Shared Processor Pool Capacity	Unattainable (Enable Shared Pool Monitoring)		10/31/2013 08:44 PM		1	5
	Free CPU Capacity	Unattainable (Enable Shared Pool Monitoring)				1	1

Figure 5-7 Error on Shared Processing Pool monitoring

To enable the feature, access the partition properties for a specific VIOS on the Hardware Management Console (HMC). On the General tab (Figure 5-8), select **Allow performance information collection**.

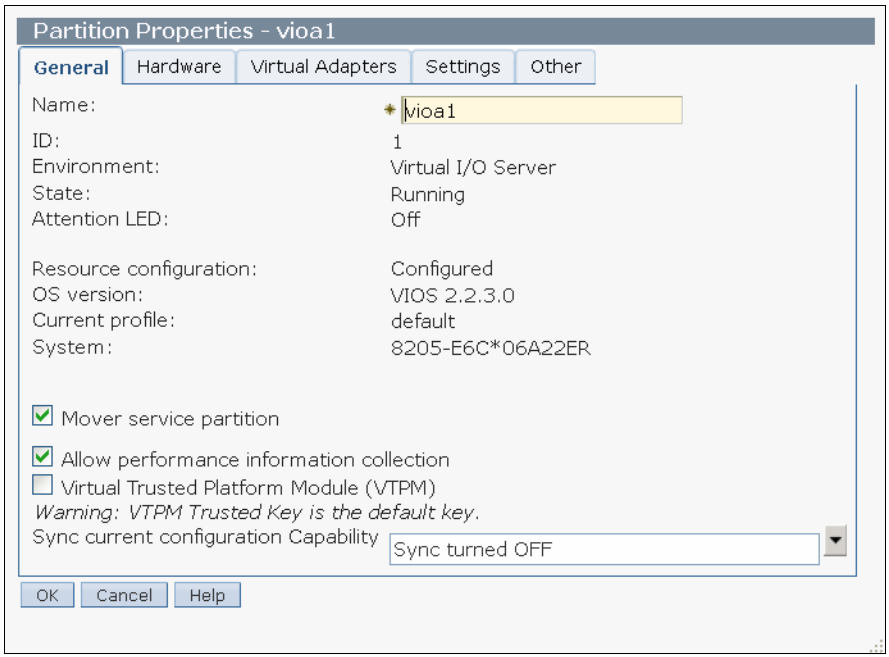


Figure 5-8 Enable Shared Processor Pool monitoring

5.3.6 Memory advisory report

The VIOS - Memory advisory report consists of the information that relates to the memory resources, such as the available free memory, paging space that is allocated, paging rate, and pinned memory. The output is similar to Figure 5-9.

VIOS - Memory — Risk/Impact 1=lowest 5=highest

	Name	Measured Value	Suggested Value	First Observed	Last Observed	Risk	Impact
✓	Real Memory	8.000 GB		10/24/2013 04:34 PM			
i	Availab Memory	5.563 GB		10/24/2013 04:34 PM			
✓	Paging Rate	0.0 MBps Paging Rate		10/24/2013 04:34 PM	10/24/2013 04:44 PM		
✓	Paging Space Size	1.500 GB		10/24/2013 04:34 PM	10/24/2013 04:44 PM		
i	Free Paging Space	1.488 GB free		10/24/2013 04:34 PM			
✓	Pinnec Memory	1.328 GB pinned		10/24/2013 04:34 PM			

Figure 5-9 Memory advisory report

5.3.7 I/O activities (disk and network) advisory report

The VIOS - I/O Activity disk adapter advisory report consists of information that relates to disk I/O activity:

- Average and peak I/O operations per second
- Network I/O activity
- Average and peak inflow and outflow I/O per second

The output is similar to Figure 5-10.

VIOS - I/O Activity

Name	Value
Disk I/O Activity	Average : 121 @ 110.53 KB Peak: 1313 @ 86KB
Network I/O Activity	[Average Send: 29 @ 25.6 MBps , Average Receieve: 37 @ 20.3MBps] [Peak Send: 685 @ 103.7 MBps , Peak Receive: 692 @ 391.8MBps]

Figure 5-10 VIOS I/O Activity (disk and network) advisory report

5.3.8 Disk Drives advisory report

The VIOS - Disk Drives advisory report consists of the information that relates to the disks attached to the VIOS, such as the I/O activities that are blocked and I/O latencies. The output is similar to Figure 5-11.

VIOS - Disk Drives *Risk/Impact 1=lowest 5=highest*

	Name	Measured Value	Suggested Value	First Observed	Last Observed	Risk	Impact
	Physica Drive Count	10		10/24/2013 04:34 PM			
	I/O Operations Blocked	pass		10/24/2013 04:34 PM	10/24/2013 04:44 PM		
	Long I/O Latency	pass		10/24/2013 04:34 PM	10/24/2013 04:44 PM		

Figure 5-11 VIOS - Disk Drives advisory report

5.3.9 Disk adapter advisory report

The VIOS - Disk Adapters advisory report consists of information that relates to the Fibre Channel adapters that are connected to the VIOS. This report illustrates the information that is based on the average I/O operations per second, adapter utilization, and running speed. The output is similar to Figure 5-12 on page 90.

Also, If you have NPIV clients, you can expand the NPIV items. The following traffic-related statistics are shown:

- Average I/O per seconds
- I/Os blocked
- Traffic by individual worldwide port name (WWPN)

VIOS - Disk Adapters

Risk/Impact 1=lowest 5=highest

Name	Measured Value	Suggested Value	First Observed	Last Observed	Risk	Impact
FC Adapter Count	2		10/24/2013 04:34 PM			
FC I/O Operations per second	0 @ 0 KB		10/24/2013 04:34 PM	10/24/2013 04:44 PM		
fcs0	Average : 0 @ 0 KB		10/24/2013 04:34 PM	10/24/2013 04:44 PM		
fcs1	Average : 0 @ 0 KB		10/24/2013 04:34 PM	10/24/2013 04:44 PM		
FC Adapter Utilization	optimal					
FC Adapter Utilization (fcs0)	high:0.0% (3.2 @ 3.8K)		10/24/2013 04:34 PM	10/24/2013 04:44 PM		
FC Adapter Utilization (fcs1)	high:0.0% (3.1 @ 13.1K)		10/24/2013 04:34 PM	10/24/2013 04:44 PM		
NPIV Client Utilization - fcs1	High: 0.00 % Average: 0.00 %		10/24/2013 04:34 PM	10/24/2013 04:44 PM		
vmlnx01	Average 0 iops @ 4 KB Peak: 1 iops @ 19 KB		10/24/2013 04:34 PM	10/24/2013 04:44 PM		
NPIV Client Utilization - fcs0	High: 0.01 % Average: 0.00 %		10/24/2013 04:34 PM	10/24/2013 04:44 PM		
vmaix01	Average 0 iops @ 1 KB Peak: 3 iops @ 5 KB		10/24/2013 04:34 PM	10/24/2013 04:44 PM		
FC I/O Operations Blocked	optimal		10/24/2013 04:34 PM	10/24/2013 04:44 PM		
FC Port Speeds	running at full speed		10/24/2013 04:34 PM	10/24/2013 04:44 PM		

Figure 5-12 VIOS - Disk Adapters advisory report

5.3.10 Shared Storage Pool advisory report

The VIOS - Shared Storage Pool advisory report consists of the information that relates to the SSP utilization based on the pool size thresholds, and per client pool utilization. The output is similar to Figure 5-13.

VIOS - Shared Storage Pool — *Risk/Impact 1=lowest 5=highest*

	Name	Measured Value	Suggested Value	First Observed	Last Observed	Risk	Impact
ℹ	SSP size (SSPpool)	49.88 GB		10/24/2013 04:34 PM			
✓	SSP utilization (SSPpool)	2.30 %		10/24/2013 04:34 PM			
ℹ	ClientID: 10 (8205-E6C0206A22ER)	0.00 GB		10/24/2013 04:34 PM			

Figure 5-13 VIOS - Shared Storage Pool advisory report

5.3.11 Shared Ethernet Adapters (SEA) advisory report

The VIOS - Shared Ethernet Adapters advisory report consists of the information that relates to SEAs, such as the count and utilization. The output is similar to Figure 5-14.

VIOS - Shared Ethernet Adapters — *Risk/Impact 1=lowest 5=highest*

	Name	Measured Value	Suggested Value	First Observed	Last Observed	Risk	Impact
ℹ	SEA Adapter Count	1		10/24/2013 04:34 PM			
✗	SEA (ent4)	Mapping: Physical : (ent0),Virtual : (ent2,ent3)		10/24/2013 04:34 PM			

Figure 5-14 VIOS - Shared Ethernet Adapters (SEA) advisory report

If you expand the SEA column, the following detail is shown (Figure 5-15 on page 92):

- SEA utilization based on the physical interfaces
- Arbitrate (baudrate) and traffic information
- Per-client SEA traffic information
- The number of SEA adapters configured
- The utilization metrics, such as average send and receive
- The SEA peak send and receive

VIOS - Shared Ethernet Adapters

Risk/Impact 1=lowest 5=highest

	Name	Measured Value	Suggested Value	First Observed	Last Observed	Risk	Impact
	SEA Adapter Count	1		11/01/2013 11:27 AM			
	SEA (ent6)	Mapping: Physical : (ent0),Virtual : (ent4,ent5)		11/01/2013 11:27 AM			
	SEA LargeSend - ent6	Enabled		11/01/2013 11:27 AM			
	SEA LargeReceive - ent6	Disabled	Enabled	11/01/2013 11:27 AM		3	5
	SEA Thread - ent6	Enabled		11/01/2013 11:27 AM			
	Phy FlowControl - ent0	Enabled		11/01/2013 11:27 AM			
	Phy LargeSend - ent0	Enabled		11/01/2013 11:27 AM			
	Phy LargeReceive - ent0	Disabled	Enabled	11/01/2013 11:27 AM		2	4
	LPAR client buffer alloc. - ent0	Optimal		11/01/2013 11:27 AM			
	Virt ent4 buffer alloc.	Optimal		11/01/2013 11:27 AM			
	Virt ent5 buffer alloc.	Optimal		11/01/2013 11:27 AM			

Figure 5-15 Expanded Shared Ethernet Adapters (SEA) status

Note: Physical LargeReceive is the feature that enables coalescing received packets into a large packet before passing them to the next layer for enhanced performance. The feature is enabled by default and it is also known as *TCP segment aggregation*. The adapter hardware offload capability can complete data aggregation faster than the operating system software. This feature can help you improve receive performance.

The following commands enable the LargeReceive function for physical device `ent1`:

```
chdev -l en1 -a state=down
chdev -l ent1 -a large_receive=yes (if hardware supports)
chdev -l en1 -a state=up
```

6

PowerVM Live Partition Mobility

PowerVM Live Partition Mobility has two major enhancements:

6.1 PowerVM Server Evacuation

Use the Hardware Management Console (HMC) at Version 7 Release 7.8.0, or later, to perform a server evacuation operation. This process is used to move all migration-capable logical partitions (LPARs) from one system to another. Any upgrade or maintenance operations can be performed after all the partitions are migrated and the source system is powered off.

You can migrate all the migration-capable AIX, Linux, and IBM i partitions from the source server to the destination server by running the following command from the HMC command line:

```
migrlpar -o m -m source_server -t target_server --all
```

The command finishes silently and the virtual machines (VMs) are in the target server. If the target server has partitions that were configured before the evacuation start, the moved LPARs will coexist with the previous partitions. To roll back the LPARs to the source server, move them individually using the HMC.

Important: The following conditions apply for a partition that is considered migration capable:

- The source server must not have any inbound or outbound migration operations in progress.
- The destination server must not have any outbound migration operations in progress.
- The destination server must have enough resources to fit the logical partitions to be migrated, at least for the minimum requirements shown in their profiles.
- The HMC must be at Version 7 Release 7.8.0, or later.

To stop the migration of all the migration-capable AIX, Linux, and IBM i partitions, run the following command from the HMC command line:

```
migrlpar –o s -m source_server --all
```

This command cancels any migration currently in progress.

6.2 Settings to improve Live Partition Mobility performance

You can improve the partition mobility performance by installing the latest available firmware, HMC, and VIOS software levels on both the source and target mover service partitions.

At the time of writing this book, the latest available levels are HMC 7.7.8.0 and VIOS 2.2.1. Firmware levels relate to each server and can vary depending on the model.

Link: For more information about software and firmware levels or to download the latest codes, go to this website:

`http://www.ibm.com/support/fixcentral`

The following tables describe the VIOS Processing Unit resources that are suggested to achieve maximum throughput. These resources are in addition to the resources already assigned to the VIOS to handle the existing virtual I/O resource requirements, using a 10 Gb network adapter for partition mobility.

Table 6-1 lists the suggested resources in addition to the configured resources for the VIOSs for a single migration to achieve maximum throughput. These extra resources are only needed on the Mover Service Partitions (MSPs).

Table 6-1 VIOS requirement for a single migration

Single migration		
Server family	**POWER7**	**POWER7+**
Dedicated processing units	> 3	> 2
Virtual processors	> 3	> 2
Using 1 Gb network adapter or near 100% of utilization using 10 Gb network adapter	1	1
Required additional memory when using 10 Gb network adapter	Add 1 GB of RAM in addition to the previous requirements.	

Table 6-2 lists suggested resources in addition to the configured resources for the VIOSs for up to 16 concurrent migrations to achieve maximum throughput. These extra resources are only needed on the Mover Service Partitions. Each Mover Service Partition only supports up to eight concurrent migrations. To reach 16 concurrent migrations, you need two Mover Service Partitions on both the source and target systems.

Table 6-2 VIOS requirement for up to 16 concurrent migrations (8 for each MSP)

Up to 16 concurrent migrations - eight for each Mover Service Partition		
	POWER7	**POWER7+**
Dedicated processing units	4	3
Virtual processors	4	3
Using 1 Gb network adapter or near 100% of utilization using 10 Gb network adapter	1	1
Required additional memory when using 10 Gb network adapter	Add 1 GB of RAM in addition to the previous requirements.	

In addition to these suggested settings for memory and processing units, there are other settings for dedicated network adapters. These other settings do not apply to virtual network adapters in the VMs, but they can be applied to the VIOSs. Follow these steps:

1. Enable the *Large Send Offload* and *Large Receive Offload* options on all network devices that are involved in partition mobility. This setting is enabled, by default, on all network adapters that support this feature. If you need to enable this setting, manually run these commands on the VIOS (Example 6-1).

Example 6-1 Enabling Large Send Offload and Large Receive Offload

```
lsdev -dev ent0 -attr | grep large
large_receive   no              Enable receive TCP segment aggregation                  True
large_send      no              Enable transmit TCP segmentation offload                True


chdev -dev ent0 -attr large_send=yes
ent0 changed

chdev -dev ent0 -attr large_receive=yes
ent0 changed


lsdev -dev ent0 -attr | grep large
large_receive   yes             Enable receive TCP segment aggregation                  True
large_send      yes             Enable transmit TCP segmentation offload                True
```

2. Set the **tcp_sendspace = 524288** and **tcp_recspace = 524288** parameters in the tunables file by running these commands on the VIOS (Example 6-2).

Example 6-2 Setting sendspace and recspace tunable parameters

```
optimizenet -list | grep space | grep tcp
tcp_recvspace             16K    16K    16K    4K     8E-1   byte              C
tcp_sendspace             16K    16K    16K    4K     8E-1   byte              C

optimizenet -perm -set tcp_recvspace=524288
Setting tcp_recvspace to 524288
Setting tcp_recvspace to 524288 in nextboot file
Change to tunable tcp_recvspace, will only be effective for future connections

optimizenet -perm -set tcp_sendspace=524288
Setting tcp_sendspace to 524288
Setting tcp_sendspace to 524288 in nextboot file
Change to tunable tcp_sendspace, will only be effective for future connections

optimizenet -list | grep space | grep tcp
tcp_recvspace             512K   16K    512K   4K     8E-1   byte              C
tcp_sendspace             512K   16K    512K   4K     8E-1   byte              C
```

3. Enable the *Jumbo Frames* option if it is supported by the environment by running these commands on the VIOS (Example 6-3).

Example 6-3 Enabling Jumbo Frames

```
lsdev -dev ent0 -attr | grep jumbo
jumbo_frames    no            Enable jumbo frames support                       True

chdev -dev ent0 -attr jumbo_frames=yes
ent0 changed

lsdev -dev ent0 -attr | grep jumbo
jumbo_frames    yes           Enable jumbo frames support                       True
```

Abbreviations and acronyms

ABI	application binary interface
AC	Alternating Current
ACL	access control list
AFPA	Adaptive Fast Path Architecture
AIO	Asynchronous I/O
AIX	Advanced Interactive Executive
APAR	authorized program analysis report
API	application programming interface
ARP	Address Resolution Protocol
ASMI	Advanced System Management Interface
BFF	Backup File Format
BIND	Berkeley Internet Name Domain
BIST	Built-In Self-Test
BLV	Boot Logical Volume
BOOTP	Bootstrap Protocol
BOS	Base Operating System
BSD	Berkeley Software Distribution
CA	certificate authority
CAA	Cluster Aware AIX
CATE	Certified Advanced Technical Expert
CD	compact disc
CD-R	compact disc recordable
CD-ROM	compact-disc read-only memory
CDE	Common Desktop Environment
CEC	central electronics complex
CHRP	Common Hardware Reference Platform
CLI	command-line interface
CLIs	command line interfaces
CLVM	Concurrent LVM
COD	Capacity Upgrade on Demand
CPU	central processing unit
CRC	cyclic redundancy check
CSM	Cluster Systems Management
CUoD	Capacity Upgrade on Demand
CVUT	Czech Technical University
DCM	Dual Chip Module
DES	Data Encryption Standard
DGD	Dead Gateway Detection
DHCP	Dynamic Host Configuration Protocol
DLPAR	dynamic LPAR
DMA	direct memory access
DMS	Deadman switch
DNS	Domain Name System
DR	dynamic reconfiguration
DRM	dynamic reconfiguration manager
DVD	digital versatile disc
DWO	Dynamic Workload Optimizer
EC	EtherChannel
ECC	error correction code
EOF	end-of-file
EPOW	emergency power-off warning
ERRM	Event Response resource manager

ESS	IBM Enterprise Storage Server®
F/C	Feature Code
FC	Fibre Channel
FC_AL	Fibre Channel Arbitrated Loop
FDX	Full Duplex
FLOP	Floating Point Operation
FRU	field-replaceable unit
FTP	File Transfer Protocol
GDPS®	IBM Geographically Dispersed Parallel Sysplex™
GID	group ID
GPFS™	IBM General Parallel File System
GUI	Graphical User Interface
HACMP™	IBM High Availability Cluster Multiprocessing
HBA	host bus adapter
HMC	Hardware Management Console
HTML	Hypertext Markup Language
HTTP	Hypertext Transfer Protocol
Hz	hertz
I/O	input/output
IBM	International Business Machines
ID	identifier
IDE	Integrated Device Electronics
IEEE	Institute of Electrical and Electronics Engineers
IP	Internet Protocol
IPAT	IP address takeover
IPL	initial program load
IPMP	IP Multipathing
ISV	independent software vendor
ITSO	International Technical Support Organization
IVM	Integrated Virtualization Manager
JFS	journaled file system
L1	level 1
L2	level 2
L3	level 3
LA	Link Aggregation
LACP	Link Aggregation Control Protocol
LAN	local area network
LDAP	Lightweight Directory Access Protocol
LED	light-emitting diode
LMB	Logical Memory Block
LPAR	logical partition
LPARs	least one of the partitions
LPP	licensed program product
LU	logical unit
LUN	logical unit number
LUs	logical units
LV	logical volume
LVCB	Logical Volume Control Block
LVM	Logical Volume Manager
MAC	Media Access Control
MBps	megabytes per second
MCM	multiple chip module
ML	Maintenance Level
MP	Multiprocessor
MPIO	Multipath I/O
MTU	maximum transmission unit
Mbps	megabits per second
NA	not available
NFS	Network File System
NIB	Network Interface Backup
NIM	Network Installation Management
NIMOL	NIM on Linux

NPIV	N_Port Identifier Virtualization
NVRAM	nonvolatile random access memory
N_PORT	Node Port
ODM	Object Data Manager
OS	operating system
OSPF	Open Shortest Path First
PCI	Peripheral Component Interconnect
PCI Express	iPeripheral Component Interconnect Express
PEX	Performance Explorer
PIC	Pool Idle Count
PID	process ID
PKI	public key infrastructure
PLM	Partition Load Manager
PMU	Performance Monitoring Unit
POST	power-on self-test
POWER	Performance Optimization with Enhanced Risc (Architecture)
PPC	Physical Processor Consumption
PPFC	Physical Processor Fraction Consumed
PTF	program temporary fix
PTX	Performance Toolbox
PURR	Processor Utilization Resource Register
PV	physical volume
PVID	Port Virtual LAN Identifier
PrU	Processing Units
QoS	quality of service
RAID	Redundant Array of Independent Disks
RAM	random access memory
RAS	reliability, availability, and serviceability
RBAC	role-based access control
RCP	Remote Copy
RDAC	Redundant Disk Array Controller
RHEL	Red Hat Enterprise Linux
RIO	remote input/output
RIP	Routing Information Protocol
RISC	reduced instruction-set computer
RMC	Resource Monitoring and Control
RPC	Remote Procedure Call
RPL	Remote Program Loader
RPM	Red Hat Package Manager
RSA	Rivest-Shamir-Adleman algorithm
RSCT	Reliable Scalable Cluster Technology
RSH	Remote Shell
SAN	storage area network
SCSI	Small Computer System Interface
SDD	Subsystem Device Driver
SDDPCM	Subsystem Device Driver Path Control Module
SEA	Shared Ethernet Adapter
SMIT	System Management Interface Tool
SMP	symmetric multiprocessor
SMS	system management services
SMT	simultaneous multithreading
SP	Service Processor
SPOT	Shared Product Object Tree
SPP	Shared Processing Pool
SRC	System Resource Controller
SRN	service request number
SSA	Serial Storage Architecture
SSH	Secure Shell

SSL	Secure Sockets Layer
SSP	Shared Storage Pool
SUID	Set User ID
SVC	SAN Volume Controller
SWMA	Software Maintenance agreement
TCP/IP	Transmission Control Protocol/Internet Protocol
TL	Technology Level
TSA	IBM Tivoli® System Automation
UDF	Universal Disk Format
UDID	Universal Disk Identification
VG	volume group
VGDA	Volume Group Descriptor Area
VGSA	Volume Group Status Area
VIOS	Virtual I/O Server
VIPA	virtual IP address
VLAN	virtual local area network

Related publications

The publications listed in this section are considered particularly suitable for a more detailed discussion of the topics covered in this book.

IBM Redbooks

For information about ordering these publications, see “Help from IBM” on page 108. Note that some of the documents referenced here might be available in softcopy only.

- *IBM Power Systems HMC Implementation and Usage Guide*, SG24-7491
- *Integrated Virtualization Manager for IBM Power Systems Servers*, REDP-4061
- *IBM PowerVM Best Practices*, SG24-8062
- *IBM PowerVM Virtualization Introduction and Configuration*, SG24-7940
- *IBM PowerVM Virtualization Managing and Monitoring*, SG24-7590
- *IBM Systems Director VMControl Implementation Guide on IBM Power Systems*, SG24-7829
- *Power Systems Memory Deduplication*, REDP-4827
- *PowerVM Migration from Physical to Virtual Storage*, SG24-7825
- *IBM PowerVM Virtualization Active Memory Sharing*, REDP-4470
- *A Practical Guide for Resource Monitoring and Control (RMC)*, SG24-6615

Other publications

These publications are also relevant as further information sources:

- The following types of documentation are located on the Internet at this website:

  `http://publib.boulder.ibm.com/infocenter/powersys/v3r1m5/index.jsp`

  - User guides
  - System management guides
  - Application programmer guides

- All command reference volumes
- File reference
- Technical reference volumes used by application programmers

- Detailed documentation about the PowerVM feature and the Virtual I/O Server is available at this website:

  http://bit.ly/1sUGhNY

- *IBM eServer iSeries Performance Tools for iSeries*, SC41-5340
- *IBM Tivoli Usage and Accounting Manager Data Collectors for UNIX and Linux User's Guide,* SC32-1556

Online resources

These websites are also relevant as further information sources:

- For further details of supported machine model types, visit this website:

  https://www-304.ibm.com/webapp/set2/sas/f/power5cm/power7.html

- The latest configuration file for a Power enterprise pool is available on the IBM Capacity on Demand website:

  http://www-03.ibm.com/systems/power/hardware/cod/offerings.html

- IBM Business Partners

  http://www.ibm.com/services/econfig/announce/index.html

- IBM internal website

  http://w3-03.ibm.com/transform/worksmart/docs/e-config.html

- Further configuration examples and details are provided in the Power Systems Information Center under the POWER7 Systems section:

  http://pic.dhe.ibm.com/infocenter/powersys/v3r1m5

- Upgrade from previous version (**updateios** command) and update packages can be downloaded from IBM Fix Central:

  http://www-933.ibm.com/support/fixcentral

- For storage hardware that is supported in VIOS, see this website:

  http://bit.ly/1oIPmLS

- VIOS information:

  http://bit.ly/1nctYzk

- VIOS Performance Advisor:

  http://www-01.ibm.com/support/docview.wss?uid=aixtools159f1226
- Capacity upgrade on demand (COD) activations:

  http://www-03.ibm.com/systems/power/hardware/cod/activations.html
- COD SSP information:

  http://www-03.ibm.com/systems/power/hardware/cod/offerings.html
- Hardware Management Console (HMC) interaction script:

  http://www.the-welters.com/professional/scripts/hmcMenu.txt
- IBM Redbooks publications:

  http://www.redbooks.ibm.com/
- IBM Systems information center: Power Systems Virtual I/O Server and Integrated Virtualization Manager commands:

  http://bit.ly/RUvXcU
- IBM System Planning Tool:

  http://www-947.ibm.com/systems/support/tools/systemplanningtool/
- IBM wikis:

  http://www.ibm.com/developerworks/wikis/dashboard.action

  – AIX Wiki - Performance Monitoring Documentation:

    http://bit.ly/1pd2WVp

  – nmon analyzer tool:

    http://bit.ly/Tum73O
- Virtual I/O Server monitoring wiki:

  http://bit.ly/1jsuC7S
- The nmon tool:

  http://www.ibm.com/developerworks/wikis/display/WikiPtype/nmon
- IBM Systems Information Centers:

  http://publib.boulder.ibm.com/eserver/?tocNode=int_17
- PowerVM:

  http://bit.ly/1qVbaGl
- Architecting for power management: The IBM POWER7 approach:

  http://bit.ly/1o6e2xV

- SSD and Powerpath information:

  http://bit.ly/1jDqyTR
- IBM Tivoli Monitoring Information Center:

  http://bit.ly/1o6e8Wm

Help from IBM

IBM Support and downloads

ibm.com/support

IBM Global Services

ibm.com/services

(0.2”spine)
0.17”<->0.473”
90<->249 pages

IBM PowerVM Enhancements
What is New in 2013

PowerVP and mobile CoD activations explained

Shared Storage Pool enhancements explained

Power Integrated Facility for Linux described

IBM Power Systems servers coupled with IBM PowerVM technology are designed to help clients build a dynamic infrastructure, helping to reduce costs, manage risk, and improve service levels.

IBM PowerVM delivers industrial-strength virtualization for IBM AIX, IBM i, and Linux environments on IBM POWER processor-based systems. IBM PowerVM V2.2.3 is enhanced to continue its leadership in cloud computing environments. Throughout the chapters of this publication, you will learn about the following topics:

- New management and performance tuning software products for PowerVM solutions. Virtual I/O Server (VIOS) Performance Advisor has been enhanced to provide support for N_Port Identifier Virtualization (NPIV) and Fibre Channel, Virtual Networking and Shared Ethernet Adapter, and Shared Storage Pool configurations. IBM Power Virtualization Performance (PowerVP) is introduced as a new visual performance monitoring tool for Power Systems servers.
- The scalability, reliability, and performance enhancements introduced with the latest versions of the VIOS, IBM PowerVM Live Partition Mobility, and the Hardware Management Console (HMC). As an example, this book goes through the Shared Storage Pool improvements that include mirroring of the storage pool, dynamic contraction of the storage pool, dynamic disk growth within the storage pool, and scaling improvements.

This book is intended for experienced IBM PowerVM users who want to enable 2013 IBM PowerVM virtualization enhancements for Power Systems. It is intended to be used as a companion to the following publications:

- *IBM PowerVM Virtualization Introduction and Configuration*, SG24-7940
- *IBM PowerVM Virtualization Managing and Monitoring*, SG24-7590

INTERNATIONAL TECHNICAL SUPPORT ORGANIZATION

BUILDING TECHNICAL INFORMATION BASED ON PRACTICAL EXPERIENCE

IBM Redbooks are developed by the IBM International Technical Support Organization. Experts from IBM, Customers and Partners from around the world create timely technical information based on realistic scenarios. Specific recommendations are provided to help you implement IT solutions more effectively in your environment.

For more information:
ibm.com/redbooks

SG24-8198-00 ISBN 0738439460