Redbooks

ibm.com/redbooks

Introduction to IBM PowerVM

Turgut Genc

Ivaylo Bozhinov

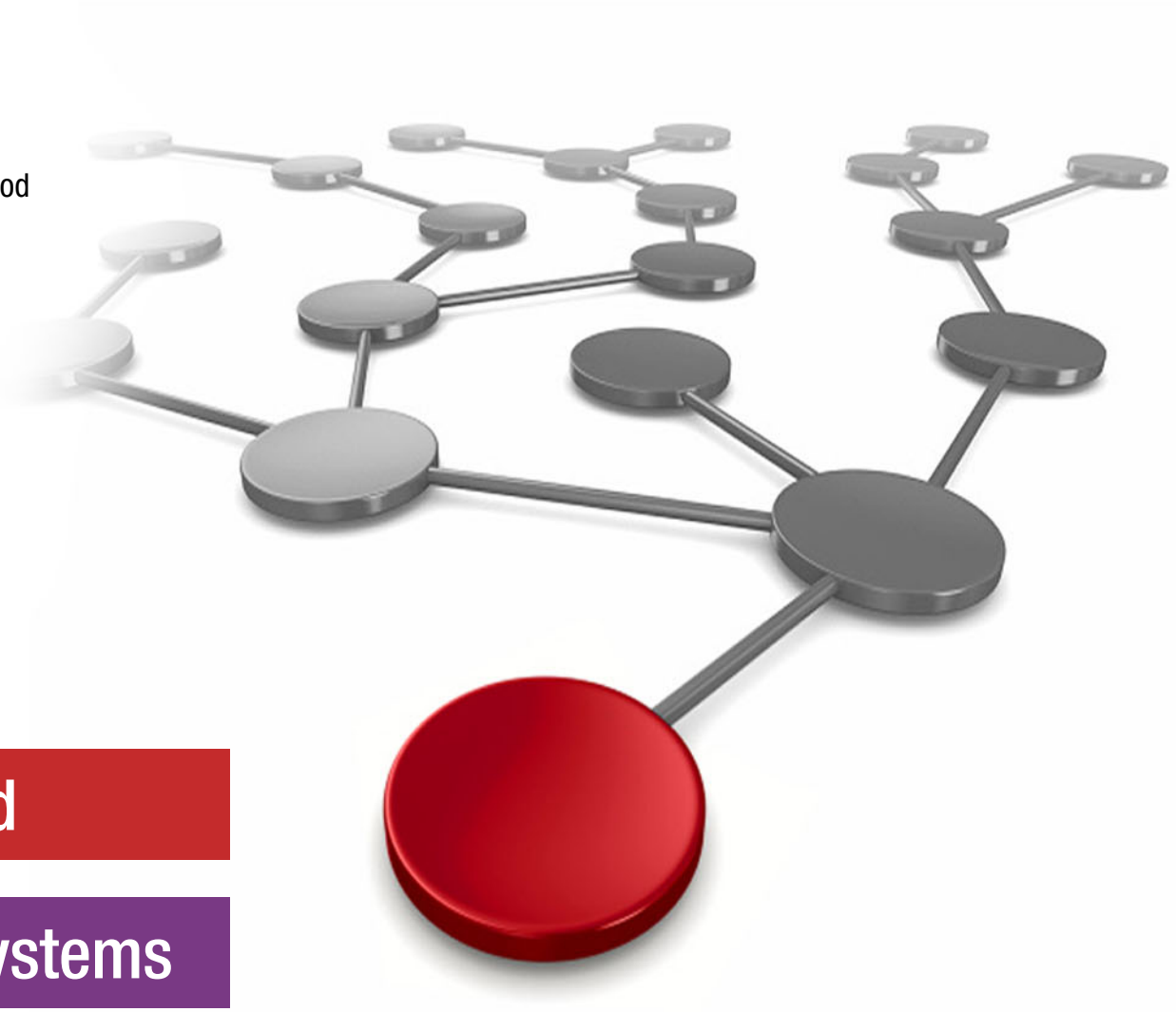
Muhammad Mahmood

Ahmed Mashhour

Ayman Mostafa

Vivek Shukla

Prerna Upmanyu



Cloud

Power Systems

IBM

Redbooks

IBM Redbooks

Introduction to IBM PowerVM

March 2023

SG24-8535-00

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

First Edition (March 2023)

This edition applies to the following versions:
- Version 7, Release 2 of IBM AIX
- Version 7, Release 4 of IBM i
- Version 3, Release 1, Service Pack (SP) 3, Fix Pack 21 of the Virtual I/O Server (VIOS)
- Version 10, Release 2, SP 1030 of the Hardware Management Console (HMC)
- Release VH950, SP 111 of the IBM Power9 System Firmware
- Release MH1010, SP 146 of the IBM Power10 System Firmware

Contents

Notices

This information was developed for products and services offered in the US. This material might be available from IBM in other languages. However, you may be required to own a copy of the product or product version in that language in order to access it.

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, MD-NC119, Armonk, NY 10504-1785, US

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 jurisdictions 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 provide in any way it believes appropriate without incurring any obligation to you.

The performance data and client examples cited are presented for illustrative purposes only. Actual performance results may vary depending on specific configurations and operating conditions.

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.

Statements regarding IBM's future direction or intent are subject to change or withdrawal without notice, and represent goals and objectives only.

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 actual people or business enterprises 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. The sample programs are provided "AS IS", without warranty of any kind. IBM shall not be liable for any damages arising out of your use of the sample programs.

Trademarks

IBM, the IBM logo, and ibm.com are trademarks or registered trademarks of International Business Machines Corporation, registered in many jurisdictions worldwide. Other product and service names might be trademarks of IBM or other companies. A current list of IBM trademarks is available on the web at “Copyright and trademark information” at http://www.ibm.com/legal/copytrade.shtml

The following terms are trademarks or registered trademarks of International Business Machines Corporation, and might also be trademarks or registered trademarks in other countries.

AIX®
Db2®
DS8000®
GDPS®
HyperSwap®
IBM®
IBM Cloud®
IBM Cloud Pak®
IBM Spectrum®
Micro-Partitioning®
OS/400®
Parallel Sysplex®
POWER®
POWER8®
POWER9™
PowerHA®
PowerVM®
Redbooks®
Redbooks (logo)  ®
SystemMirror®
Tivoli®
WebSphere®

The following terms are trademarks of other companies:

The registered trademark Linux® is used pursuant to a sublicense from the Linux Foundation, the exclusive licensee of Linus Torvalds, owner of the mark on a worldwide basis.

Java, and all Java-based trademarks and logos are trademarks or registered trademarks of Oracle and/or its affiliates.

Red Hat, Ansible, and OpenShift are trademarks or registered trademarks of Red Hat, Inc. or its subsidiaries in the United States and other countries.

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

VMware, and the VMware logo are registered trademarks or trademarks of VMware, Inc. or its subsidiaries in the United States and/or other jurisdictions.

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

Preface

Virtualization plays an important role in resource efficiency by optimizing performance, reducing costs, and improving business continuity. IBM PowerVM® provides a secure and scalable server virtualization environment for IBM AIX®, IBM® i, and Linux applications. PowerVM is built on the advanced reliability, availability, and serviceability (RAS) features and leading performance of IBM Power servers.

This IBM Redbooks® publication introduces PowerVM virtualization technologies on Power servers. This publication targets clients who are new to Power servers and introduces the available capabilities of the PowerVM platform. This publication includes the following chapters:

- Chapter 1, “IBM PowerVM overview” on page 1 introduces PowerVM and provides a high-level overview of the capabilities and benefits of the platform.
- Chapter 2, “IBM PowerVM features in details” on page 31 provides a more in-depth review of PowerVM capabilities for system administrators and architects to familiarize themselves with its features.
- Chapter 3, “Planning for IBM PowerVM” on page 77 provides planning guidance about PowerVM to prepare for the implementation of the solution.
- Chapter 4, “Implementing IBM PowerVM” on page 139 describes and details configuration steps to implement PowerVM, starting from implementing the Virtual I/O Server (VIOS) to storage and network I/O virtualization configurations.
- Chapter 5, “Managing the IBM PowerVM environment” on page 181 focuses on systems management, day-to-day operations, monitoring, and maintenance.
- Chapter 6, “Automation on IBM Power servers” on page 207 explains available techniques, utilities, and benefits of modern automation solutions.

Authors

This book was produced by a team of specialists from around the world.

Turgut Genc is a Senior Consultant at IBM Technology Services (FKA Lab Services) in the UK. He holds an MSc degree in Computer Science Engineering from Yildiz Technical University in Istanbul. He is a versatilist with 19 years of experience on IBM Power server and AIX. Turgut is the Power to Cloud Rewards EMEA team leader for IBM Power Virtual Server (PowerVS), Enterprise Pools, Performance, Migration workshops, and Technology Services tools.

Ivaylo Bozhinov is a Technical Support Professional for the IBM Power hardware division in Sofia, Bulgaria. He is a subject matter expert (SME) with a focus on service processors, enterprise Baseboard Management Controller (eBMC), Flexible Service Processor (FSP), and PowerVM hypervisor (PHYP). Ivaylo joined IBM in 2015 and contributed to numerous educational and client-related workshops and presentations. He holds a bachelor's degree in Information Technology from the State University of Library and Information Technology and a master's degree in Cybersecurity from New Bulgarian University. He supports many clients from the banking and telecom industries and retail sector.

Muhammad Farrukh Mahmood is a Senior IT Managing Consultant in Pakistan. He works for IBM Technology Services Middle East and Pakistan. He has 19 years of experience, and for the last 12 years has been working for IBM Power server and IBM Storage Systems. His areas of expertise include AIX, system migration, storage data migration, PowerVM, BigFix, and AIX security. He supports many customers in the banking and the telecom sectors.

Ahmed Mashhour is a Power Technology Services Consultant Lead at IBM Egypt. He is an IBM L2 certified Expert. He holds IBM AIX, Linux, and IBM Tivoli® certifications. He has 17 years of professional experience in IBM AIX and Linux systems. He is an IBM AIX back-end SME who supports several customers in the US, Europe, and the Middle East. His core expertise is in IBM AIX, Linux systems, clustering management, AIX security, virtualization tools, and various IBM Tivoli and database products. He authored several publications inside and outside IBM, including co-authoring other IBM Redbooks publications. He also hosted IBM AIX, Security, PowerVM, IBM PowerHA®, and IBM Spectrum® Scale classes worldwide.

Ayman Mostafa is an IBM PowerVM and VIOS Software Engineer and a team leader for the IBM PowerVM Product Support team that is based in Cairo, Egypt. Ayman joined IBM in 2015 after earning a bachelor's degree in Electronics and Communication Engineering from the Higher Institute of Engineering, El Shorouk Academy in 2007. Ayman's passion for virtualization and automation technologies led him to build a broad and deep knowledge of the key virtualization platform from IBM. Ayman is recognized as a technical leader by his team and IBM clients.

Vivek Shukla is a presales consultant for cloud, AI, and cognitive offerings in India, and he is an IBM Certified L2 (Expert) Technical Specialist. He has over 20 years of IT experience in infrastructure consulting, AIX, and IBM Power servers and storage implementations. He also has hands-on experience with IBM Power servers, AIX and system software installations, request for proposal (RFP) understandings, statement of work (SOW) preparations, sizing, performance tuning, root cause analysis (RCA), disaster recovery (DR), and mitigation planning. He wrote several Power server FAQs and is a worldwide focal point for Techline FAQs Flash. He holds a master’s degree in Information Technology from IASE University and a bachelor's degree (BTech) in Electronics and Telecommunication Engineering from IETE, New Delhi. His areas of expertise include Red Hat OpenShift, Cloud Paks, and hybrid cloud.

Prerna Upmanyu is a Software Performance Analyst in the Cognitive Systems Power Servers performance team in India. She holds an M.Tech degree in Software Systems from BITS Pilani. Prerna has over 15 years of experience working with customers designing and deploying solutions on IBM Power server. She focuses on areas such as Automation and Data Lakes-based design and deployments. Prerna’s areas of expertise include system performance, availability, and automation.

Special thanks to *Turgut Genc* for his unique contributions to this publication and the overall project. Turgut was the technical leader for this project. He provided technical guidance to the team throughout the entire writing and reviewing process.

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

Chris Engel, Austin Fowler, George Gaylord, Daniel Goldener, Chuck Graham, Pete Heyrman, Stuart Jacobs, Bob Kovacs, Dave Larson, Vani Ramagiri, Adrian Robinson, Ken Vossen
IBM US

Hari Muralidharan
IBM India

Stuart Cunliffe and Richard Moult
IBM UK

Lars Johannesson
IBM Denmark

Elvis Metodiev and Dimitar Mitev
IBM Bulgaria

Mohamed Badr and Mohamed El-Tayeb
IBM Egypt

Martin Keen and Upendra Rajan
IBM Systems Technical Training

Marcela Adan and Wade Wallace
IBM Redbooks

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 IBM Redbooks 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, IBM Redbooks
  Dept. HYTD Mail Station P099
  2455 South Road
  Poughkeepsie, NY 12601-5400

Stay connected to IBM Redbooks

- Find 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

IBM PowerVM overview

This chapter introduces PowerVM and provides a high-level overview of the IBM Power servers and PowerVM capabilities.

This chapter covers the following topics:

- PowerVM introduction
- IBM Power
- PowerVM facts and features
- PowerVM resiliency and availability
- PowerVM scalability
- PowerVM security
- PowerVM and the cloud
- PowerVC enhanced benefits

1.1 PowerVM introduction

PowerVM is an enterprise-class virtualization solution that provides a secure, flexible, and scalable virtualization for Power servers. PowerVM enables logical partitions (LPARs) and server consolidation. Clients can run AIX, IBM i, and Linux operating systems on Power servers with a world-class reliability, high availability (HA), and serviceability capabilities together with the leading performance of the Power platform.

This solution provides workload consolidation that helps clients control costs and improves overall performance, availability, flexibility, and energy efficiency. Power servers, which are combined with PowerVM technology, help consolidate and simplify your IT environment. Key capabilities include the following ones:

- Improve server utilization and I/O resource-sharing to reduce total cost of ownership and better use IT assets.
- Improve business responsiveness and operational speed by dynamically reallocating resources to applications as needed to better match changing business needs or handle unexpected changes in demand.
- Simplify IT infrastructure management by making workloads independent of hardware resources so that you make business-driven policies to deliver resources based on time, cost, and service-level requirements.

PowerVM is a combination of hardware enablement and added value to software. PowerVM consists of these major components:

- IBM PowerVM hypervisor (PHYP)
- Virtual I/O Server (VIOS)
- Service processor, enterprise Baseboard Management Controller (eBMC), or Flexible Service Processor (FSP)
- Hardware Management Console (HMC)
- PowerVM NovaLink
- IBM Power Virtualization Center (PowerVC)

Figure 1-1 on page 3 provides an overview of a Power server with multiple virtual machines (VMs) securely accessing resources, which is facilitated by PHYP.

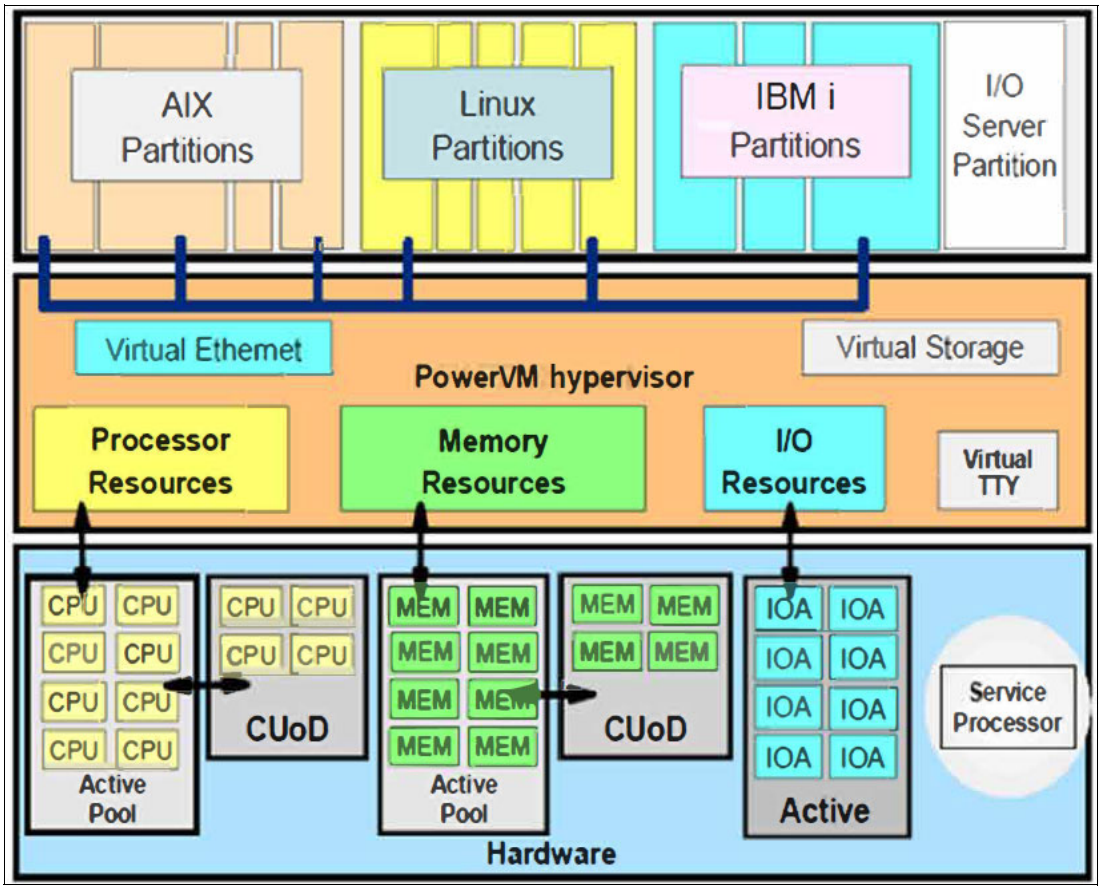


Figure 1-1 PowerVM overview

PowerVM enables virtualization of the hardware, from processor to memory and storage I/O to network I/O resources. It enables platform-level capabilities like Live Partition Mobility (LPM) or Simplified Remote Restart (SRR).

Processor virtualization

The Power servers family gives you the freedom to either use the scale-up or scale-out processing model to run the broadest selection of enterprise applications without the costs and complexity that are often associated with managing multiple physical servers. PowerVM can help eliminate underutilized servers because it is designed to pool resources and optimize their usage across multiple application environments and operating systems. Through advanced VM capabilities, a single VM can act as a separate AIX, IBM i, or Linux operating environment that uses dedicated or shared system resources. With shared resources, PowerVM can automatically adjust pooled processor or storage resources across multiple operating systems, borrowing capacity from idle VMs to handle high resource demands from other workloads.

With PowerVM on Power servers, you have the power and flexibility to address multiple system requirements in a single machine. IBM Micro-Partitioning® supports multiple VMs per processor core. Depending on the Power servers model, you can run up to 1000 VMs on a single server, each with its own processor, memory, and I/O resources. Processor resources can be assigned at a granularity of 0.01 of a core. Consolidating systems with PowerVM can help cut operational costs, improve availability, ease management, and improve service levels, and businesses can quickly deploy applications.

Multiple shared processor pools (MSPP) allow for the automatic nondisruptive balancing of processing power between VMs that are assigned to shared pools, resulting in increased throughput. MSPP also can cap the processor core resources that are used by a group of VMs to potentially reduce processor-based software licensing costs.

Shared Dedicated Capacity allows for the “donation” of spare CPU cycles from dedicated processor VMs to a shared processor pool (SPP). Because a dedicated VM maintains absolute priority for CPU cycles, enabling this feature can increase system utilization without compromising the computing power for critical workloads.

Because its core technology is built into the system firmware, PowerVM offers a highly secure virtualization platform that received the Common Criteria Evaluation and Validation Scheme (CCEVS) EAL4+ certification3 for its security capabilities.

Memory virtualization

The PHYP automatically virtualizes the memory by dividing it into standard logical memory blocks (LMBs) and manages memory assignments to partitions by using hardware page tables (HPTs). This technique enables translations from effective addresses to physical real addresses and running multiple operating systems simultaneously in their own secured logical address space. The PHYP uses some of the activated memory in a Power server to manage memory that is assigned to individual partitions, manage I/O requests, and support virtualization requests.

PowerVM also features IBM Active Memory Expansion (AME), a technology that allows the effective maximum memory capacity to be much larger than the true physical memory for AIX partitions. AME uses memory compression technology to transparently compress in-memory data, which allows more data to be placed into memory and expand the memory capacity of configured systems. The in-memory data compression is managed by the operating system, and this compression is transparent to applications and users. AME is configurable on a per-LPAR basis. Thus, AME can be selectively enabled for one or more LPARs on a system.

I/O virtualization

The VIOS is a special purpose VM that can be used to virtualize I/O resources for AIX, IBM i, and Linux VMs. VIOS owns the resources that are shared by VMs. A physical adapter that is assigned to the VIOS can be shared by many VMs, which reduces the cost by eliminating the need for dedicated I/O adapters. PowerVM provides virtual SCSI (vSCSI) and N_Port ID Virtualization (NPIV) technologies to enable direct or indirect access to storage area networks (SANs) from multiple VMs. VIOS facilitates Shared Ethernet Adapter (SEA), which is a component that bridges a physical Ethernet adapter and one or more Virtual Ethernet Adapters (VEAs).

PowerVM supports single-root I/O virtualization (SR-IOV) technology, which allows a single I/O adapter to be shared concurrently with multiple LPARs. This capability provides hardware-level speeds with no additional CPU usage because the adapter virtualization is enabled by the adapter at the hardware level.

The SR-IOV implementation on Power servers has an extra feature that is called virtual Network Interface Controllers (vNICs). A vNIC is backed by an SR-IOV logical port (LP) on the VIOS, which supports LPM of VMs that use SR-IOV.

PowerVM also offers a capability that is called Hybrid Network Virtualization (HNV). HNV allows a partition to leverage the efficiency and performance benefits of SR-IOV LPs and participate in mobility operations. HNV leverages existing technologies such as AIX Network Interface Backup (NIB), IBM i Virtual IP Address (VIPA), and Linux active-backup bonding as its foundation, and introduces new automation for configuration and mobility operations.

When we talk about PowerVM, we are referring to the components, features, and technologies that are listed in the Table 1-1 and Table 1-2.

Table 1-1 Major components of PowerVM

Components	Function provided by
PHYP	Hardware platform
VIOS	Hypervisor or VIOS
HMC	HMC
PowerVM NovaLink	Hypervisor or Novalink
Service Processor	eBMC or FSP
Virtualization Management Interface (VMI)	Hypervisor

Table 1-2 PowerVM features and technologies

Category	Features and technologies	Function provided by
Server	LPAR	Hypervisor
Server	Dynamic logical partitioning (DLPAR)	Hypervisor
Server	Capacity on Demand (CoD)	Hypervisor or HMC
Server	LPM	Hypervisor, VIOS, or HMC
Server	SRR	Hypervisor or HMC
Server	AIX Workload Partitions (WPARs)	AIX
Server	IBM System Planning Tool (SPT)	SPT
Processor Virtualization	Micro-Partitioning	Hypervisor
Processor Virtualization	Processor compatibility mode	Hypervisor
Processor Virtualization	Simultaneous multithreading (SMT)	Hardware
Processor Virtualization	SPPs	Hypervisor
Memory Virtualization	Active Memory Mirroring (AMM)	Hardware
Memory Virtualization	AME	Hardware or AIX
Storage Virtualization	Shared storage pools (SSPs)	Hypervisor or VIOS
Storage Virtualization	vSCSI	Hypervisor or VIOS
Storage Virtualization	Virtual Fibre Channel (NPIV)	Hypervisor or VIOS

Category	Features and technologies	Function provided by
Storage Virtualization	Virtual Optical Device and Tape	Hypervisor or VIOS
Network Virtualization	VEA	Hypervisor or HMC
Network Virtualization	SEA	Hypervisor, VIOS, or HMC
Network Virtualization	Single-root I/O virtualization (SR-IOV)	Hypervisor or adapter
Network Virtualization	SR-IOV with vNIC	Hypervisor, adapter, or VIOS
Network Virtualization	HNV	Hypervisor, adapter, VIOS, or HMC

Table 1-3 lists the deprecated PowerVM features and technologies.

Table 1-3 Deprecated PowerVM features

Features and technologies	Function provided by
Integrated Virtualization Manager (IVM)	Hypervisor, VIOS, or IVM
Active Memory Sharing	Hypervisor or VIOS
Active Memory Deduplication	Hypervisor
Partition Suspend/Resume	Hypervisor or VIOS
Versioned WPARs	AIX
Host Ethernet Adapter (HEA)	Hypervisor
PowerVP	Hypervisor

1.2 IBM Power

The *Power* name is closely aligned with the physical processors that implement the Power architecture. The family of servers that are based on IBM POWER® processors is collectively referred as *IBM Power*.

Power servers are the core element of the PowerVM ecosystem. Power servers are widely known for their reliability, scalability, and performance characteristics for the most demanding workloads. They provide enterprise class virtualization for resource management and flexibility. Over decades, in every generation, Power servers were continuously enhanced in every aspect (performance, reliability, scalability, and flexibility), providing a wide range of options and solutions to customers.

Figure 1-2 on page 7 showcases the IBM sustained and consistent roadmap of technology advancements with Power servers.

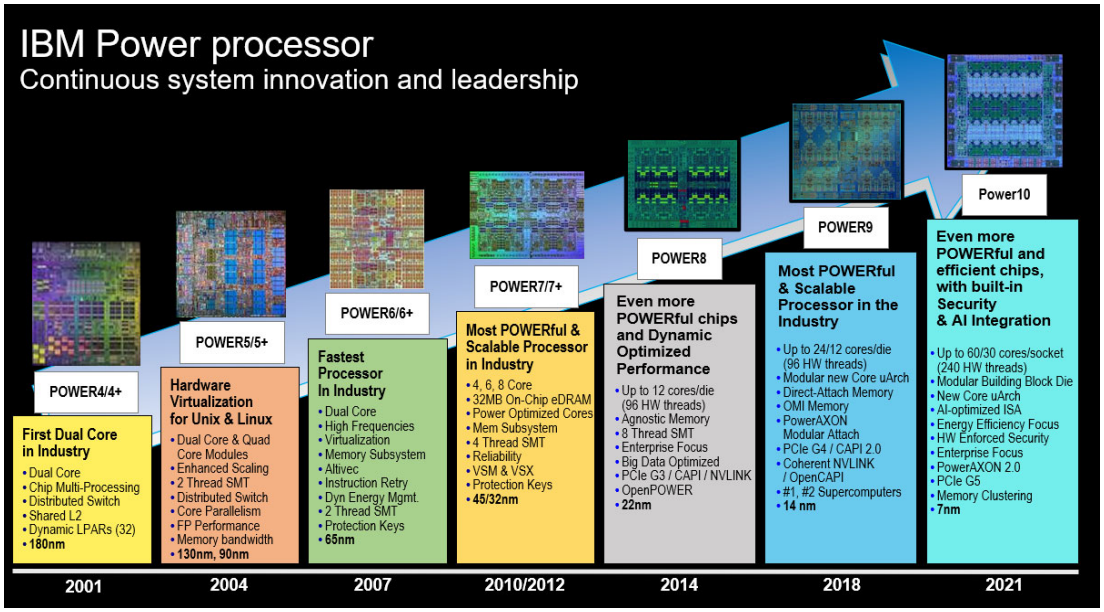


Figure 1-2 IBM POWER processor roadmap

The IBM Power family of scale-out and scale-up servers includes workload consolidation platforms that help clients control costs and improve overall performance, availability, and energy efficiency.

At the time of writing, Power10 is the latest available processor technology in the portfolio, with offerings that range from enterprise-class systems to scale-out systems. The list of available Power10 processor-based systems is as follows:

- Enterprise scale-up
  - IBM Power E1080 (9080-HEX)
  - IBM Power E1050 (9043-MRX)
- Scale-out
  - IBM Power S1024 (9105-42A)
  - IBM Power S1022 (9105-22A)
  - IBM Power S1022s (9105-22B)
  - IBM Power S1014 (9105-41B)
  - IBM Power L1014 (9786-42H)
  - IBM Power L1022 (9786-22H)
  - IBM Power L1024 (9786-42H)

For more information about Power servers, see IBM Power servers, found at:

`https://www.ibm.com/it-infrastructure/power`

All Power servers are equipped with the PHYP, which is embedded in the system firmware. When a machine is powered on, the PHYP is automatically loaded together with the system firmware. It enables virtualization capabilities at the heart of the hardware and minimizes the virtualization overhead as much as possible.

Some of the Power servers might be used as a stand-alone server without an HMC. However, the most commonly preferred approach is to use an HMC to extend capabilities of the PowerVM, including LPAR and virtualization.

Figure 1-3 provides a high-level overview of a Power server that is managed by an HMC.

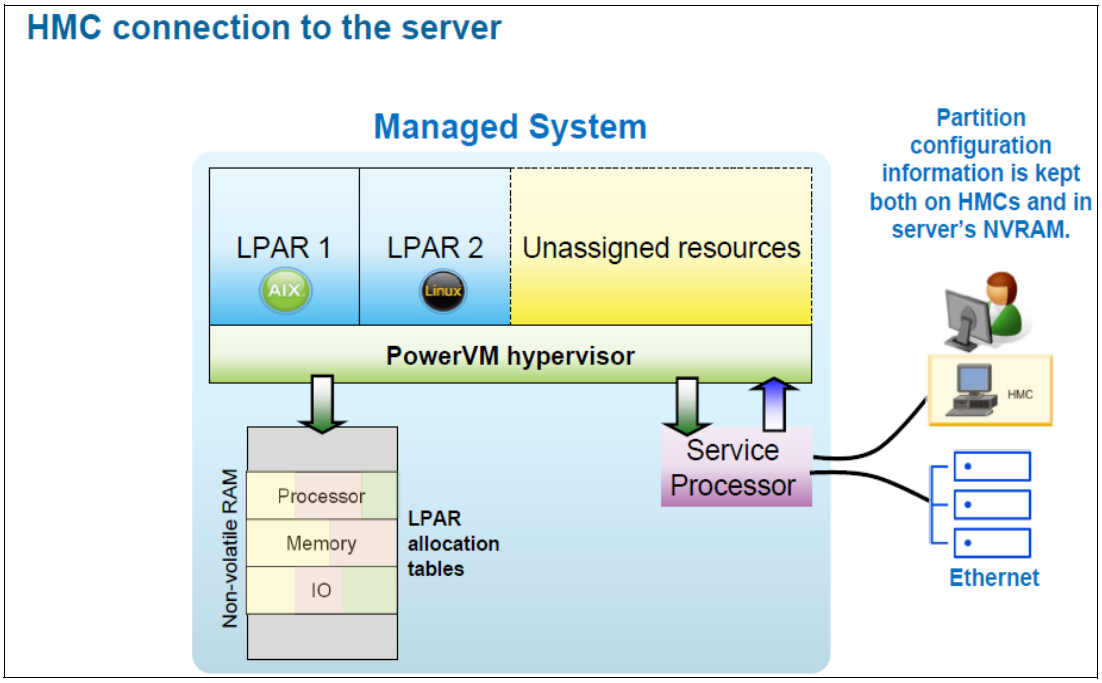


Figure 1-3 Power server that is managed by a Hardware Management Console

1.2.1 OpenPOWER

In August 2013, IBM worked closely with Google, Mellanox, Nvidia, and Tyan to establish the OpenPOWER Foundation. The goal of this collaboration is to establish an open ecosystem that is based on the IBM Power architecture.

In August 2019, IBM announced that the OpenPOWER Foundation would become part of the Linux Foundation. At the same time, IBM further opened the architecture to the world by allowing anybody to implement and manufacture their own processors capable of running software that is built for the Power Instruction Set Architecture (ISA) at no additional charge. IBM continues to base its POWER processors (including Power10) on the Power ISA, which is managed by the foundation. Since its inception, membership in the OpenPOWER foundation grew to over 150 organizations and academic partners.

1.2.2 Operating systems support on IBM Power servers

IBM Power architecture is bi-Endian, which allows running operating systems that are designed either for Big Endian or Little Endian platforms. As a result, Power servers support running IBM AIX, IBM i, and Linux workloads simultaneously.

For more information, see the following resources:

- IBM AIX Standard Edition, found at:

  https://www.ibm.com/in-en/it-infrastructure/power/os/aix
- IBM i operating system, found at:

  https://www.ibm.com/in-en/it-infrastructure/power/os/ibm-i
- Enterprise Linux servers, found at:

  https://www.ibm.com/in-en/it-infrastructure/power/os/linux

1.3 PowerVM facts and features

PowerVM is built on the Power platform and virtualization technologies that are enabled by the PHYP and the VIOS.

Note: Historically, three editions of PowerVM, which were suited for various purposes, were available. These editions were PowerVM Express Edition, PowerVM Standard Edition, and PowerVM Enterprise Edition. Today, PowerVM Enterprise Edition is the only option, and it is automatically included in all Power servers.

PowerVM Enterprise Edition consists of the following components and features:

- PHYP
- VIOS
- HMC
- PowerVM NovaLink
- Service processor
- VMI
- LPAR
- DLPAR
- CoD
- IBM Power Enterprise Pools (PEP)
- LPM
- SRR
- AIX WPAR
- SPT
- Micro-Partitioning technology
- POWER processor compatibility modes
- SMT
- SPP
- PowerVM AMM
- PowerVM AME
- SSP
- Thin provisioning
- vSCSI
- Virtual Fibre Channel (NPIV)
- Virtual optical device and tape
- VEA
- SEA
- SR-IOV
- SR-IOV with vNIC adapters
- HNV

The remainder of this chapter briefly describes these PowerVM features.

1.3.1 PowerVM hypervisor

The PHYP is the foundation of PowerVM. The PHYP divides physical system resources into isolated LPARs. Each LPAR operates like an independent server that runs its own operating system, such as AIX, IBM i, Linux, or VIOS. The PHYP can assign dedicated processors, memory, and I/O resources, which can be dynamically reconfigured as needed to each LPAR. The PHYP also can assign shared processors to each LPAR by using its Micro-Partitioning feature. The PHYP creates an SPP from which it allocates virtual processors to the LPARs as needed. In other words, the PHYP creates virtual processors so that LPARs can share the physical processors while running independent operating environments.

Combined with features of the IBM POWER processors, the PHYP delivers functions that enable capabilities such as dedicated processor partitions, Micro-Partitioning, virtual processors, IEEE virtual local area network (VLAN) compatible virtual switches, VEAs, vSCSI adapters, virtual Fibre Channel (VFC) adapters, and virtual consoles.

The PHYP is a firmware layer between the hosted operating systems and the server hardware, as shown in Figure 1-4.

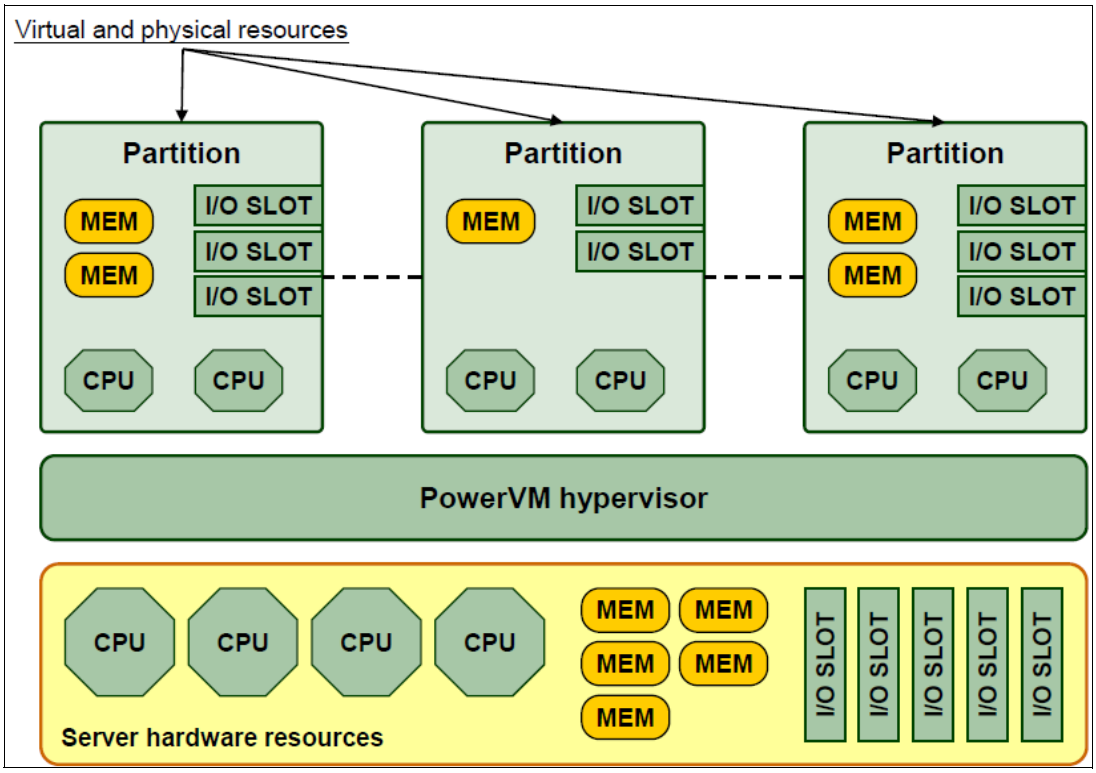


Figure 1-4 The PowerVM hypervisor abstracts physical server hardware

The PHYP is always installed and activated regardless of system configuration. The PHYP has no specific or dedicated processor resources that are assigned to it.

The PHYP performs the following tasks:

- Enforces partition integrity by providing a security layer between LPARs.
- Provides an abstraction layer between the physical hardware resources and the LPARs that use them. It controls the dispatch of virtual processors to physical processors, and saves and restores all processor state information during a virtual processor context switch.
- Controls hardware I/O interrupts and management facilities for partitions.

The PHYP firmware and the hosted operating systems communicate with each other through PHYP calls (hcalls).

1.3.2 Virtual I/O Server

As part of PowerVM, the VIOS is a software appliance with which you can associate physical resources and share these resources among multiple client LPARs.

The VIOS can provide both virtualized storage and virtualized network adapters to the virtual I/O clients. The goal is achieved by exporting the physical device on the VIOS through vSCSI. Alternatively, virtual I/O clients can access independent physical storage through the same or different physical Fibre Channel (FC) adapter by using NPIV technology. Virtual Ethernet enables IP-based communication between LPARs on the same system by using a virtual switch that is facilitated by the hypervisor that can work with VLANs.

For storage virtualization that uses vSCSI, these backing devices can be used:

- Direct-attached entire disks from the VIOS.
- SAN disks that are attached to the VIOS.
- Logical volumes that are defined on either of the previously mentioned disks.
- File-backed storage, with files that are on either of the previously mentioned disks.
- Logical units (LUs) from SSPs.
- Optical storage devices.
- Tape storage devices.

For storage that use NPIV, the VIOS facilitates VFC to FC mapping, which enables these backing devices:

- SAN disks that are directly presented to the VM, passing through VIOS.
- Tape Storage devices that are directly presented to the VM, passing through VIOS.

For virtual Ethernet, you can define SEAs on the VIOS, bridging network traffic between the server internal virtual Ethernet networks and external physical Ethernet networks.

The VIOS technology facilitates the consolidation of LAN and disk I/O resources and minimizes the number of physical adapters that are required while meeting the nonfunctional requirements of the server.

The VIOS can run in either a dedicated processor partition or a shared processor partition (Micro-Partition).

Figure 1-5 shows a basic VIOS configuration. This diagram shows only a small subset of the capabilities to illustrate the basic concept of how the VIOS works. The physical resources such as the physical Ethernet adapter and the physical disk adapter are accessed by the client partition by using virtual I/O devices.

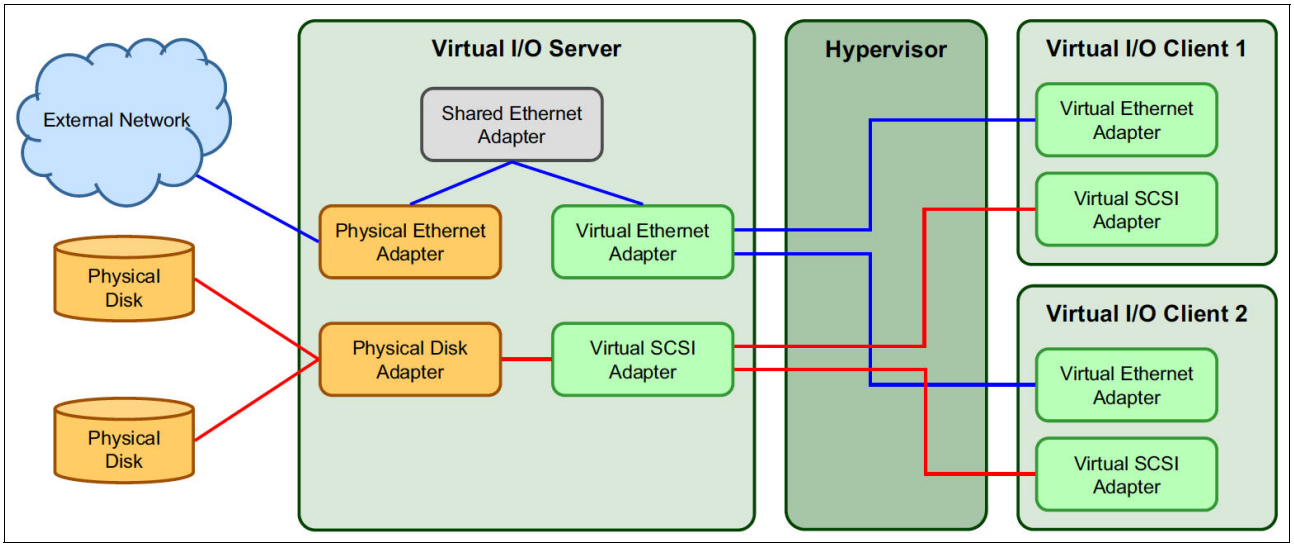


Figure 1-5 Simple Virtual I/O Server configuration

1.3.3 Hardware Management Console

The HMC is a dedicated Linux-based appliance that you use to configure and manage IBM Power servers. The HMC provides access to LPAR functions, service functions, and various system management functions through both a browser-based interface and a command-line interface (CLI). Because it is an external component, the HMC does not consume any resources from the systems that it manages, and you can maintain it without affecting system activity. The HMC can be configured for remote access either by using an Secure Shell (SSH) client or a web browser. The user can access only the HMC management application. The user is not allowed to install other applications.

A second HMC can be connected to a single Power server for redundancy purposes, which is called a dual HMC configuration. An HMC can manage up to 48 Power servers and a maximum of 2000 VMs.

The HMC enhances PowerVM with the following capabilities:

- Virtual console access for VMs
- VM configuration and operation management
- CoD management
- Service tools
- LPM
- SRR

At the time of writing, two types of HMCs are available:

- A rack-mounted physical HMC appliance
- A virtual HMC (vHMC) appliance

An HMC virtual appliance can be configured on a VM within a Power server. A Power server-based vHMC is not allowed to manage its own host Power server.

An HMC virtual appliance on x86 environment can be configured by using one of the following hypervisors:

- KVM hypervisor
- Xen hypervisor
- VMware ESXi

1.3.4 PowerVM NovaLink

PowerVM NovaLink is a software interface that is used for virtualization management. You can install PowerVM NovaLink on a Power server. PowerVM NovaLink enables highly scalable modern cloud management and deployment of critical enterprise workloads. You can use PowerVM NovaLink to provision many VMs on Power servers quickly and at a reduced cost.

PowerVM NovaLink runs on a Linux LPAR on IBM POWER8®, IBM POWER9™, or Power10 processor-based systems that are virtualized by PowerVM. You can manage the server through a Representational State Transfer application programming interface (REST API) or through a CLI. You can also manage the server by using PowerVC or other OpenStack solutions. PowerVM NovaLink is available at no additional charge for servers that are virtualized by PowerVM.

PowerVM NovaLink provides the following benefits:

- Rapidly provisions many VMs on Power servers.
- Simplifies the deployment of new systems. The PowerVM NovaLink installer creates a PowerVM NovaLink partition and VIOS partitions on the server and installs operating systems and the PowerVM NovaLink software. The PowerVM NovaLink installer reduces the installation time and facilitates repeatable deployments.
- Reduces the complexity and increases the security of your server management infrastructure. PowerVM NovaLink provides a server management interface on the server. The server management network between PowerVM NovaLink and its VMs is secure by design and is configured with minimal user intervention.
- Operates with PowerVC or other OpenStack solutions to manage your servers.

For more information, see PowerVM NovaLink, found at:

https://www.ibm.com/docs/en/power10/9080-HEX?topic=environment-powervm-novalink

1.3.5 Service processor

The *service processor* is a separate, independent processor that provides hardware initialization during system load, monitoring of environmental and error events, and maintenance support. The HMC is connected to the managed system through an Ethernet connection to the service processor. The service processor performs many vital reliability, availability, and serviceability (RAS) functions. It provides the means to diagnose, check th status of, and sense operational conditions of a remote system, even when the main processor is inoperable. The service processor enables firmware and operating system surveillance; several remote power controls; environmental monitoring; reset and boot features; remote maintenance; and diagnostic activities.

Up to the Power E1080 model, every Power server is equipped with FSPs. Starting with Power10 scale-out and midrange (Power E1050) models, servers now are configured with the eBMC service processor instead of the FSP.

1.3.6 Virtualization Management Interface

The eBMC-managed Power servers have a new model to communicate between the HMC and the PHYP in addition to a connection to the Baseboard Management Controller (BMC). This interface is called the VMI. Administrators must configure the VMI IP address after an eBMC-based system is connected to the HMC. As a best practice, configure the VMI IP before powering on the system. The HMC communicates with two endpoints to manage the system: the BMC itself and VMI.

Figure 1-6 shows the difference between an FSP-based system and an eBMC-based system. The VMI interface shares the physical connection of the eBMC port but has its own IP address.

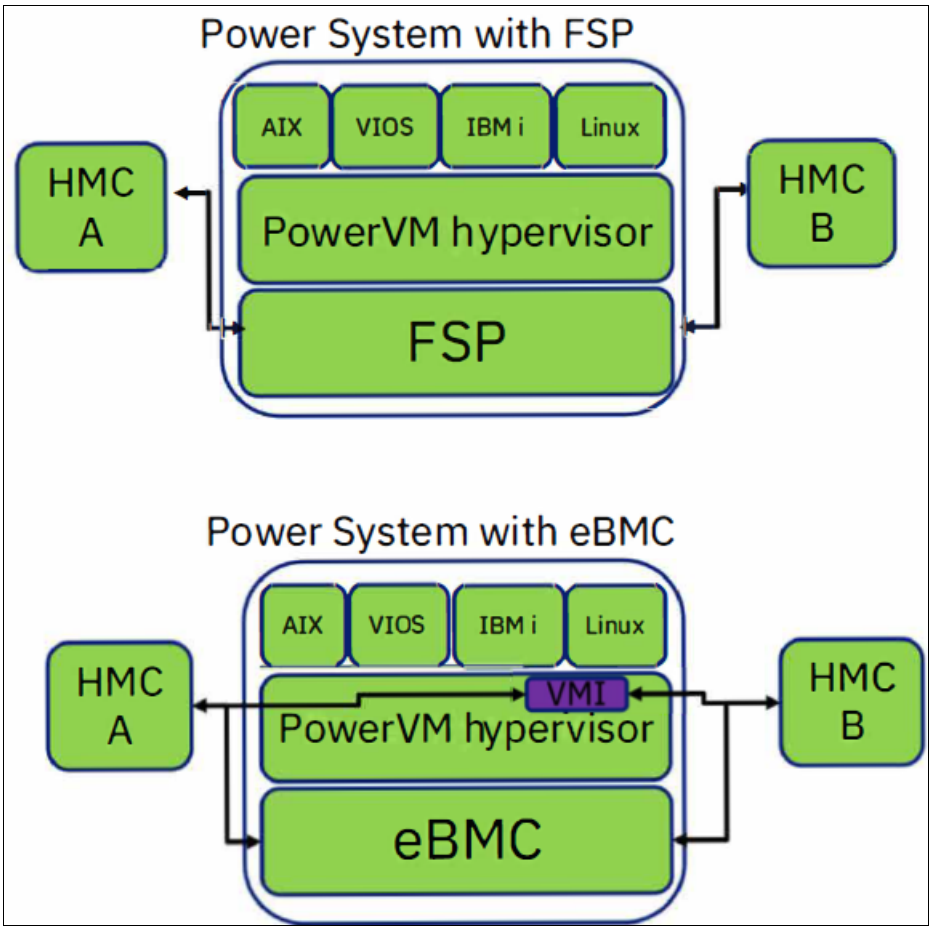


Figure 1-6 Comparing an FSP-based system with an eBMC-based system

1.3.7 Logical partitioning

Logical partitioning is the ability to make a server run as though it were two or more independent servers. When you logically partition a server, you divide the resources on the server into subsets called LPARs. You can install software on an LPAR, and the LPAR runs as an independent logical server with the resources that you allocated to the LPAR.

You can assign processors, memory, and input/output devices to LPARs. You can run AIX, IBM i, Linux, and the VIOS in LPARs. The VIOS provides virtual I/O resources to other LPARs with general-purpose operating systems.

Note: The terms LPAR and VM have the same meaning and are used interchangeably in this publication.

LPARs share a few system attributes, such as the system serial number, system model, and processor feature code. All other system attributes can vary from one LPAR to another one.

You can create a maximum of 1000 LPARs on a server. You must use tools to create LPARs on your servers. The tool that you use to create LPARs on each server depends on the server model and the operating systems and features that you want to use on the server.

Each LPAR has its own of the following components:

- Operating system
- Licensed Internal Code (LIC) and open firmware
- Console
- Resources
- Other items that are expected in a stand-alone operating system environment, such as:
  - Problem logs
  - Data (libraries, objects, and file systems)
  - Performance characteristics
  - Network identity
  - Date and time

Each LPAR can be:

- Dynamically modified
- Relocated (LPM)

Benefits of using partitions

Some benefits that are associated with the usage of partitions are as follows:

- Capacity management.

  Flexibility in allocating system resources
- Consolidation:
  - Consolidate multiple workloads that are running on different hardware and software licenses; reduce floor space; support contracts; and use in-house support and operations.
  - Efficient use of resources. Dynamically reallocate resources and LPM to support performance and availability.
- Application isolation on a single frame:
  - Separate workloads.
  - Guaranteed resources.
  - Data integrity.

- Merge production and test environments.

  Test on the same hardware on which you deploy the production environment.

For more information about LPARs, see Logical partition overview, found at:

https://www.ibm.com/docs/en/power10/9080-HEX?topic=partitioning-logical-partition-overview

1.3.8 Dynamic logical partitioning

DLPAR refers to the ability to move resources between partitions without shutting down the partitions. This goal can be accomplished from the HMC application or by using the HMC CLI. These resources include the following ones:

- Processors, memory, and I/O slots.
- The ability to add and remove virtual devices.

DLPAR operations do not weaken the security or isolation between LPARs. A partition sees only resources that are explicitly allocated to the partition along with any potential connectors for more virtual resources that might be configured. Resources are reset when moved from one partition to another one. Processors are reinitialized; memory regions are cleared; and adapter slots are reset.

DLPAR operations depend on Resource Monitoring and Control (RMC) communication in between the HMC and the LPAR.

DLPAR is described in more detail in 2.5, “Dynamic logical partitioning” on page 52.

For more information about DLPAR, see Dynamic logical partitioning, found at:

https://www.ibm.com/docs/en/aix/7.3?topic=management-dynamic-logical-partitioning

1.3.9 Capacity on Demand

With CoD offerings, you can dynamically activate one or more resources on your server as your business peaks dictate. You can activate inactive processor cores or memory units that are installed on your server on a temporary and permanent basis. CoD offerings are available on selected IBM servers.

CoD is described in more detail in 2.9, “Capacity on Demand” on page 66.

For more information about CoD, see *Capacity on Demand*, found at:

https://www.ibm.com/docs/en/power10/9080-HEX?topic=environment-capacity-demand

1.3.10 Power Enterprise Pools

PEP are infrastructure licensing models that allow resource flexibility and sharing among Power servers, which enables cost efficiency. PEP are built on the CoD capability of Power servers, which allows consuming processor and memory resources beyond the initial configuration of pool member machines. Two different PEP are available:

- PEP 1.0
- PEP 2.0 (Power Systems Private Cloud with Shared Utility Capacity)

IBM PEP 1.0 is built on mobile capacity for processor and memory resources. You can move mobile resource activations among the systems in a pool with HMC commands. These operations provide flexibility when you manage large workloads in a pool of systems and helps to rebalance the resources to respond to business needs. This feature is useful for providing continuous application availability during maintenance. Workloads, processor activations, and memory activations can be moved to other systems. Disaster recovery (DR) planning also is more manageable and cost-efficient because of the ability to move activations where and when they are required.

IBM PEP 2.0, also known as Power Systems Private Cloud with Shared Utility Capacity, provides enhanced multisystem resource sharing and by-the-minute consumption of on-premises compute resources to clients who deploy and manage a private cloud infrastructure. All installed processors and memory on servers in a Power Enterprise Pool 2.0 are activated and made available for immediate use when a pool is started. Processor and memory usage on each server are tracked by the minute and aggregated across the pool by IBM Cloud® Management Console (IBM CMC). Base Processor Activation features and Base Memory Activation features and the corresponding software license entitlements are purchased for each server in a Power Enterprise Pool 2.0.

The base resources are aggregated and shared across the pool without having to move them from server to server. The unpurchased capacity in the pool can be used on a pay-as-you-go basis. Resource usage that exceeds the pool's aggregated base resources is charged as metered capacity by the minute and debited against purchased capacity credits on a real-time basis. Capacity credits can be purchased from IBM, an authorized IBM Business Partner, or online through the IBM Entitled Systems Support (IBM ESS) website, where available.

Shared Utility Capacity simplifies system management, so clients can focus on optimizing their business results instead of moving resources and applications around within their data center. Resources are tracked and monitored by IBM CMC, which automatically tracks usage by the minute and debits against Capacity Credits, which are based on actual usage. With Shared Utility Capacity, you no longer need to worry about over-provisioning capacity to support growth because all resources are activated on all systems in a pool. Purchased Base Activations can be seamlessly shared between systems in a pool, and all unpurchased capacity can be used on a pay-per-use basis.

PEP are described in more detail in 2.10, “Power Enterprise Pools” on page 70.

IBM Cloud Management Console

IBM CMC for IBM Power servers provides a consolidated view of the Power servers in your enterprise. It runs as a service that is hosted in IBM Cloud, and you can access it securely anytime and anywhere to monitor and gain insights about your Power servers. IBM CMC can be deployed based on the client input to different IBM Cloud regions. Dynamic views of performance, inventory, and logging for a complete Power enterprise, whether on-premises or off-premises, simplifies and unifies information in a single location. This information allows clients to easily make more informed decisions. As private and hybrid cloud deployments grow, enterprises need new insight into these environments. Tools that provide consolidated information and analytics can be key enablers to smooth operation of infrastructure.

The following applications are available on IBM CMC:

- Inventory
- Logging
- Patch Planning
- Capacity Monitoring
- Enterprise Pools 2.0

1.3.11 Live Partition Mobility

By using *partition mobility*, a component of the PowerVM Enterprise Edition hardware feature, you can migrate AIX, IBM i, and Linux LPARs from one system to another one. The mobility process transfers the system environment, which includes the processor state, memory, attached virtual devices, and connected users.

By using active partition migration (LPM), you can migrate AIX, IBM i, and Linux LPARs that are running, including the operating system and applications, from one system to another one. The LPAR and the applications that are running on that migrated LPAR do not need to be shut down.

By using inactive partition migration, you can migrate a powered-off AIX, IBM i, or Linux LPAR from one system to another one.

You can use the HMC to migrate an active or inactive LPAR from one server to another one.

LPM is described in more detail in 2.6, “Partition mobility” on page 54.

For more information about LPM, see Partition mobility, found at:

https://www.ibm.com/docs/en/power10/9080-HEX?topic=environment-live-partition-mobility

1.3.12 Simplified Remote Restart

SRR is a HA option for LPARs. When an error causes a server outage, a partition that is configured with the SRR capability can be restarted on a different physical server. Sometimes, it might take longer to start the server, in which case the SRR feature can be used for faster reprovisioning of the partition. This operation completes faster compared to restarting the server that failed and then restarting the partition.

The SRR feature is supported on IBM Power8 or later processor-based systems.

Here are the characteristics of the SRR feature:

- During the SRR operation, the LPAR is shut down and then restarted on a different system.
- The SRR feature preserves the resource configuration of the partition. If processors, memory or I/O are added or removed while the partition is running, the SRR operation activates the partition with the most recent configuration.

The SRR feature is not supported from the HMC for LPARs that are co-managed by the HMC and PowerVM NovaLink. However, you can run SRR operations by using PowerVC with PowerVM NovaLink.

SRR is described in more detail in 2.7, “Simplified Remote Restart” on page 59.

For more information, see Simplified Remote Restart, found at:

https://www.ibm.com/docs/en/power10/9786-42H?topic=9786-42H/p10eew/p10eew_remmres.html

1.3.13 AIX Workload Partitions

WPARs are virtualized operating system environments within a single instance of the AIX operating system. WPARs secure and isolate the environment for the processes and signals that are used by enterprise applications.

For more information about WPARs, see IBM Workload Partitions for AIX, found at:

https://www.ibm.com/docs/en/aix/7.2?topic=workload-partitions-aix

1.3.14 IBM System Planning Tool

The SPT is a browser-based application that helps you design system configurations. It is useful for designing LPARs. You can use SPT to plan a system that is based on existing performance data or based on new workloads. System plans that are generated by the SPT can be deployed on the system by the HMC. The SPT is available to help the user with system planning, design, and validation, and to provide a system validation report that reflects the user's system requirements while not exceeding system recommendations.

For more information about SPT, see IBM System Planning Tool for POWER processor-based systems, found at:

https://www.ibm.com/support/pages/ibm-system-planning-tool-power-processor-based-systems-0

1.3.15 Micro-Partitioning

Micro-Partitioning technology enables the configuration of multiple partitions to share system processing power. All processors that are not dedicated to specific partitions are placed in the SPP that is managed by the hypervisor. Partitions that are set to use shared processors can use the SPP. You can set a partition that uses shared processors to use as little as 0.05 processing units, which is approximately a 20th of the processing capacity of a single processor. You can specify the number of processing units to be used by a shared processor partition down to a 100th of a processing unit. This ability to assign fractions of processing units to partitions and allowing partitions to share processing units is called *Micro-Partitioning* technology.

For more information about Micro-Partitioning, see Micro-Partitioning technology, found at:

https://www.ibm.com/docs/en/power10/9080-HEX?topic=powervm-micro-partitioning-technology

1.3.16 POWER processor compatibility modes

Processor compatibility modes enable you to migrate LPARs between servers that have different processor types without upgrading the operating environments that are installed in the LPARs.

You can run several versions of the AIX, IBM i, Linux, and VIOS operating environments in LPARs. Sometimes earlier versions of these operating environments do not support the capabilities that are available with new processors, which limit your flexibility to migrate LPARs between servers that have different processor types.

A processor compatibility mode is a value that is assigned to an LPAR by the hypervisor that specifies the processor environment in which the LPAR can successfully operate. When you migrate an LPAR to a destination server that has a different processor type from the source server, the processor compatibility mode enables that LPAR to run in a processor environment on the destination server in which it can successfully operate. In other words, the processor compatibility mode enables the destination server to provide the LPAR with a subset of processor capabilities that are supported by the operating environment that is installed in the LPAR.

For more information about processor compatibility modes, see Processor compatibility modes, found at:

https://www.ibm.com/docs/en/power10/9080-HEX?topic=mobility-processor-compatibility-modes

1.3.17 Simultaneous multithreading

SMT is a processor technology that allows multiple instruction streams (threads) to run concurrently on the same physical processor, which improves overall throughput. The principle behind SMT is to allow instructions from more than one thread to be run concurrently on a processor. This capability allows the processor to continue performing useful work even if one thread must wait for data to be loaded. To the operating system, each hardware thread is treated as an independent logical processor. Single-threaded (ST) execution mode is also supported. Power8 or later processors support eight SMT threads per core. Different VMs on the same server might be configured with different SMT values on the run time.

Figure 1-7 shows the relationship between physical, virtual, and logical processors.

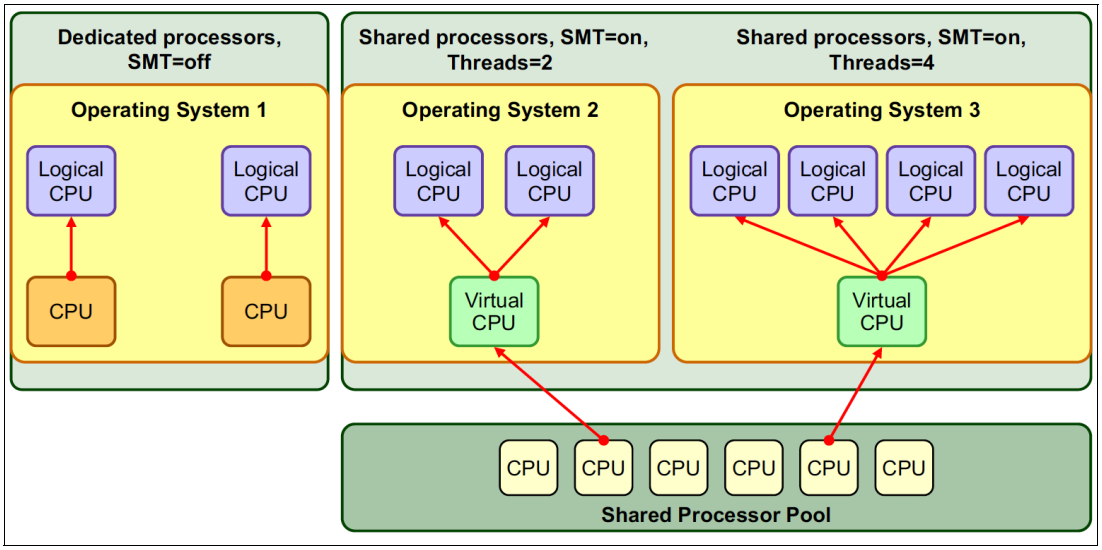


Figure 1-7 Physical, virtual, and logical processors

1.3.18 Shared processor pools

An SPP is a hypervisor assisted capability, which allows a specific group of Micro-Partitions (and their associated virtual processors) to share physical processing resources.

Shared processors, SPP, and MSPP are described in detail in 2.1.3, “Shared processors” on page 34 and 2.1.5, “Multiple shared processor pools” on page 36.

1.3.19 Active Memory Mirroring

AMM for the hypervisor is a RAS feature that is designed to ensure that system operation continues even if the unlikely event of an uncorrectable error occurs in the main memory that is used by the system hypervisor. When this feature is activated, two identical copies of the system hypervisor are maintained in memory. Both copies are simultaneously updated with any changes. This feature is also sometimes referred to as *system firmware mirroring*.

Note: AMM does not mirror partition data. It mirrors only the hypervisor code and its components, allowing this data to be protected against a DIMM failure.

If a memory failure on the primary copy occurs, the second copy is automatically called, which eliminates platform outages due to uncorrectable errors in system hypervisor memory. The hypervisor code LMBs are mirrored on distinct DDIMMs to enable more usable memory. No specific DDIMM hosts the hypervisor memory blocks, so the mirroring is done at the LMB level, not at the DDIMM level. To enable the AMM feature, the server must have enough free memory to accommodate the mirrored memory blocks. The SPT can help to estimate the amount of memory that is required.

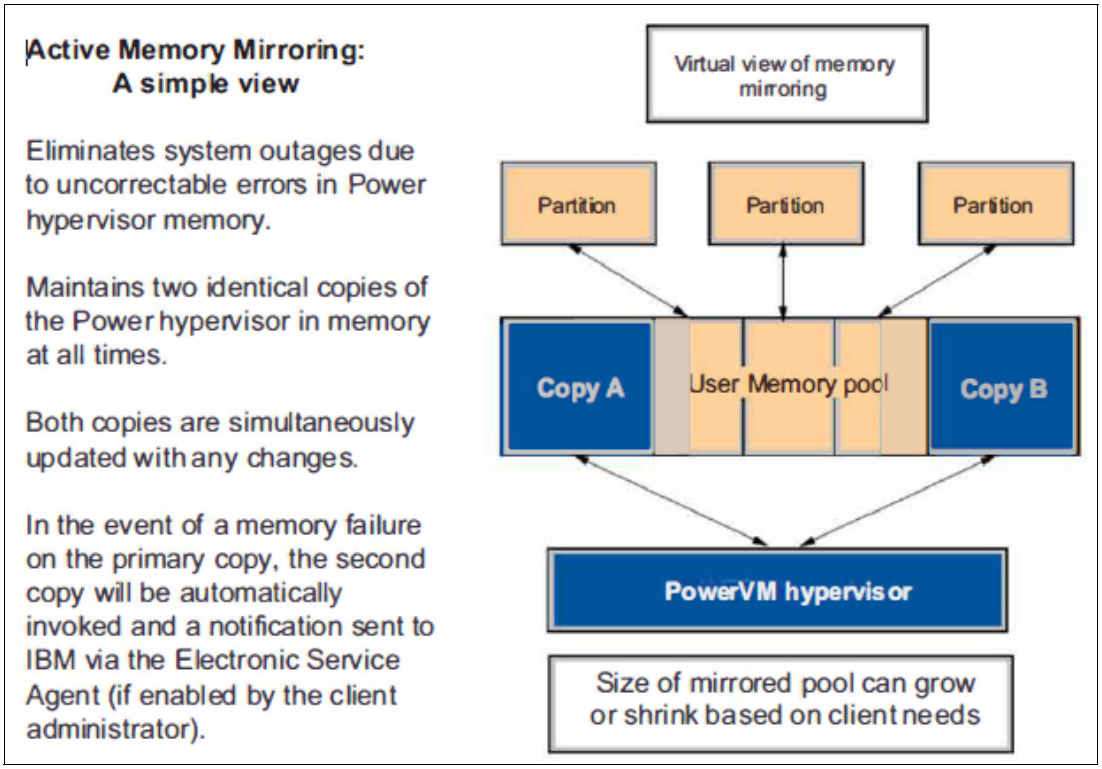


Figure 1-8 A simple view of Active Memory Mirroring

1.3.20 Active Memory Expansion

AME is an AIX OS feature. It enables in-flight memory compression and decompression by using hardware accelerators. The objective is to achieve better memory utilization and efficiency through a tradeoff between more processing capacity and more usable memory for the applications.

AME is described in more detail in 2.2.2, “Active Memory Expansion” on page 39.

1.3.21 Shared storage pools

An SSP is a server-based storage virtualization that is clustered. It is an extension of existing storage virtualization on the VIOS. In this feature, SAN-based LUNs are represented to a cluster of VIOSs and pooled together to allow VIOS administrators to create LUs out of the shared pool.

SSP is described in more detail in 2.3.3, “Shared storage pools” on page 44.

1.3.22 Virtual SCSI

vSCSI refers to a virtualized implementation of the SCSI protocol. vSCSI is based on a client/server relationship. The VIOS owns the physical resources and acts as server, or in SCSI terms, a target device. The client LPARs access the vSCSI backing storage devices that are provided by the VIOS as clients.

vSCSI is described in more detail in 2.3.1, “Virtual SCSI” on page 42.

1.3.23 Virtual Fibre Channel

NPIV is an industry-standard technology that allows an NPIV-capable FC adapter to be configured with multiple virtual worldwide port names (WWPNs). This technology also is called *VFC*. Similar to the vSCSI function, VFC is another way of securely sharing a physical FC adapter among multiple VIOS client partitions.

NPIV is described in more detail in 2.3.2, “Virtual Fibre Channel” on page 43.

1.3.24 Virtual optical device and tape

A CD or DVD device or a tape device that is attached to a VIOS can be virtualized and assigned to virtual I/O clients.

1.3.25 Virtual Ethernet Adapters

Virtual Ethernet allows the administrator to define in-memory connections between partitions that are handled at the system level (PHYP and operating systems interaction). These connections are represented as VEAs and exhibit characteristics like physical high-bandwidth Ethernet adapters. They support the industry standard protocols, such as IPv4, IPv6, ICMP, or Address Resolution Protocol (ARP).

VEAs are described in more detail in 2.4.1, “Virtual Ethernet Adapter” on page 46.

1.3.26 Shared Ethernet Adapter

A SEA is a VIOS component that bridges a real Ethernet adapter and one or more VEAs. A SEA allows many client partitions to effectively share physical network resources and communicate with networks outside of the server.

SEA is described in more detail in 2.4.2, “Shared Ethernet Adapter” on page 47.

1.3.27 Single-root I/O virtualization

SR-IOV is an extension to the Peripheral Component Interconnect Express (PCIe) specification that allows a single I/O adapter to be shared concurrently with multiple LPARs. It provides hardware level speeds with no additional CPU usage because the adapter virtualization is enabled by the adapter at the hardware level. However, this performance comes at the cost of the loss of LPM and SRR capabilities.

SR-IOV with vNIC

SR-IOV with vNIC is the virtualized version of the SR-IOV technology. It enables LPM between servers and allows up to six backing devices for hypervisor-assisted automated failover. Since vNIC virtualization is established by using VIOSs, more CPU is used on the VIOS.

SR-IOV is described in more detail in 2.4.3, “Single-root I/O virtualization” on page 47.

1.3.28 Hybrid Network Virtualization

HNV allows AIX, IBM i, and Linux partitions to leverage the efficiency and performance benefits of SR-IOV LPs and participate in mobility operations such as LPM and SRR. HNV is enabled by selecting a new **Migratable** option when an SR-IOV LP is configured. HNV uses active backup bonding to allow LPM for the partitions that are configured with an SR-IOV LP.

HNV is described in more detail in 2.4.5, “Hybrid Network Virtualization” on page 51.

1.4 PowerVM resiliency and availability

Power servers with PowerVM come with industry-leading resiliency features that are embedded into its layers, plus software solutions to solidify the availability and robustness of the entire ecosystem.

Starting with the POWER processor and server architecture, Power servers are equipped with redundant components and techniques that are used to provide outstanding resiliency and availability.

PowerVM resiliency starts in the hardware layer. Examples include redundancy of power and cooling components, and service processors and system clocks.

In the hypervisor layer, PowerVM offers AMM capability to keep two copies of the system memory.

In the management layer, PowerVM supports dual HMC and also supports HMC and Novalink for the same server concurrently.

In the virtualization layer, PowerVM supports multiple pairs of VIOSs. Redundant VIOSs (commonly referred as dual VIOS) are recommended for group of workloads. Dual VIOS enables higher throughput and availability, and allows resiliency through failovers in between active VIOSs for storage and network virtualization.

In the architectural level, PowerVM is equipped and surrounded with software solutions, which offer higher levels of availability by applying industry best practices. These software solutions include IBM PowerHA SystemMirror®, IBM VM Recovery Manager (VMRM) HA, and IBM VM Recovery Manager for DR.

VMRM is described in more detail in 2.8, “VM Recovery Manager” on page 62.

To learn more about Power server RAS features, see Introduction to IBM Power Reliability, Availability, and Serviceability for IBM POWER9 processor-based systems by using IBM PowerVM, found at:

https://www.ibm.com/downloads/cas/2RJYYJML

1.5 PowerVM scalability

PowerVM supports up to 16 VIOSs, with 20 partitions per processor and 1000 LPARs per server. Each partition requires a minimum of one I/O slot for disk attachment and a second I/O slot for Ethernet attachment. Therefore, at least 2000 I/O slots are required to support 1000 LPARs when dedicated physical adapters are used, and that is before any resilience or adapter redundancy is considered.

The high-end IBM Power servers can provide many physical I/O slots by attaching expansion drawers, but it is nowhere near to the number of required physical I/O slots mentioned earlier. Also, the mid-end servers have a lower maximum number of I/O ports than high-end servers. To overcome these physical requirements, I/O resources are typically shared. vSCSI and VFC technologies, facilitated by the VIOS, provide the means to share I/O resources for storage. Also, virtual Ethernet, SEA, and SR-IOV based technologies provide the means for network resource sharing.

IBM Power servers can scale up to four system nodes, four processor sockets, and 16 TB of memory per system node, totaling a maximum of 64 TB memory, 16 processor sockets, and 240 cores in Power10 processor-based systems.

1.6 PowerVM security

Security is one of the prime concerns for any digital organization today. Persistent end-to-end security is needed to reduce the exposure to potential security threats. PowerVM offers a platform with industry-leading security features for your workloads by closely coupling with built-in hardware-enabled security capabilities of Power servers.

The key PowerVM security features are as follows:

- Workload isolation

  PowerVM provides data isolation between the deployed partitions.

- Secure Boot

  A secure initial program load (IPL) process or the Secure Boot feature allows only correctly signed firmware components to run on the system processors. Each component of the firmware stack, including hostboot, the PHYP, and partition firmware, is signed by the platform manufacturer and verified as part of the IPL process.

- OS Secure Boot (Linux and AIX)

  The OS Secure Boot feature extends the chain of trust to the LPAR by digitally verifying the OS boot loader, kernel, runtime environment, device drivers, kernel extensions, applications, and libraries.

- Trusted Platform Module (TPM)

  A framework to support remote attestation of the system firmware stack through a hardware TPM.

- Virtual Trusted Platform Module (vTPM)

  You can enable a vTPM on an LPAR by using the HMC after the LPAR is created.

- Platform KeyStore (PKS)

  An AES-256 GCM encrypted non-volatile store to provide LPARs with more capabilities to protect sensitive information. PowerVM provides an isolated PKS storage allocation for each partition with individually managed access controls. Some of the possible use cases of this feature include the following ones:

  - Boot device encryption.
  - Self-encrypting drives
  - Unlocking encrypted logical volumes without requiring a passphrase.
  - Public key and certificate protection.
  - Provide a lockable flash that is accessible during an early IPL of the partition that is then locked down from further access.

- Transparent Memory Encryption (TME)

  The Power10 family of servers introduces a new layer of defense with end-to-end memory encryption. All data in memory remains encrypted while in transit between memory and processor. Because this capability is enabled at the silicon level, there is no extra management setup and performance impact.

- Fully Homomorphic Encryption (FHE)

  Power10 processor-based servers include four times more crypto engines in every core compared to Power9 processor-based servers to accelerate encryption performance across the stack. These innovations, along with new in-core defense for return-oriented programming attacks and support for post-quantum encryption and FHE, makes one of the most secure server platforms even better.

For more information about Secure Boot in PowerVM, see Secure Boot in PowerVM, found at:

https://www.ibm.com/docs/en/power10/9080-HEX?topic=powervm-secure-boot-in

For more information about Secure Boot in AIX, see Secure boot, found at:

https://www.ibm.com/docs/en/aix/7.2?topic=configuration-secure-boot

For more information about vTPM, see Creating a logical partition with Virtual Trusted Platform capability, found at:

https://www.ibm.com/docs/en/power10/9080-HEX?topic=partitions-creating-logical-partition-virtual-trusted-platform-capability

For more information about how to secure the PowerVM environment, see Security, found at:

https://www.ibm.com/docs/en/power10/9043-MRX?topic=information-security

For more information about Platform KeyStore, see Enabling the platform keystore capability on a logical partition, found at:

https://www.ibm.com/docs/en/power10/9080-HEX?topic=partitions-enabling-platform-keystore-capability-logical-partition

For more information about TME, see 2.1.9, “Pervasive memory encryption”, in *IBM Power E1080 Technical Overview and Introduction*, REDP-5649.

For more information about FHE, see Homomorphic Encryption Services, found at:

https://www.ibm.com/security/services/homomorphic-encryption

Power servers benefit from the integrated security management capabilities that are offered by IBM PowerSC for managing security and compliance on Power servers (AIX, IBM i, and Linux on Power). PowerSC introduces extra features to help customers manage security end-to-end for virtualized environments that are hosted on Power servers. For more information, see IBM PowerSC, found at:

https://www.ibm.com/products/powersc

Figure 1-9 depicts the security features and capabilities that the Power servers stack provides for applications and workloads that are deployed on Power.

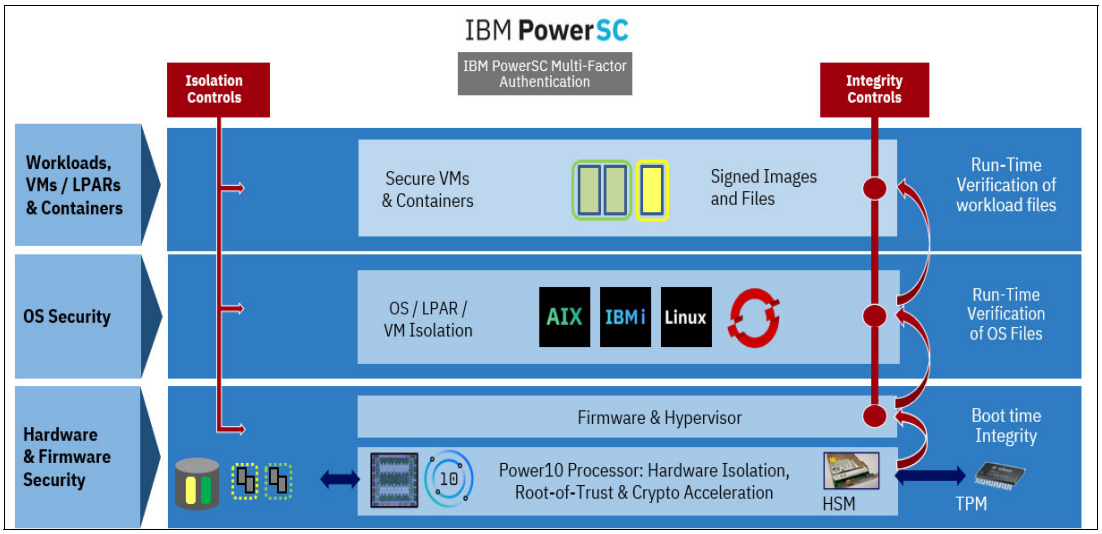


Figure 1-9 PowerVM Security

1.7 PowerVM and the cloud

This section describes IBM PowerVM in a cloud environment.

IBM PowerVC provides simplified virtualization management and cloud deployments for AIX, IBM i, and Linux VMs running on IBM Power servers. PowerVC is designed to build private cloud capabilities on Power servers and improve administrator productivity. It can further integrate with cloud environments through higher-level cloud orchestrators.

IBM Power Systems Virtual Server (PowerVS) is a Power offering. The Power Systems Virtual Servers are in the IBM data centers, distinct from the IBM Cloud servers with separate networks and direct-attached storage. The environment is in its own pod and the internal networks are fenced but offer connectivity options to meet customer requirements.

This infrastructure design enables PowerVS to maintain key enterprise software certification and support because the PowerVS architecture is identical to the certified on-premises infrastructure. The virtual servers, also known as LPARs, run on IBM Power server hardware with the PHYP.

Either in a Power Private Cloud with IBM PowerVC or in a public cloud like IBM PowerVS, PowerVM plays a key role when you implement Power servers in the cloud environment.

For more information, see the following resources:

- IBM PowerVM, found at:

  https://www.ibm.com/products/ibm-powervm
- Advanced virtualization and cloud management, found at:

  https://www.ibm.com/products/powervc
- *What is a Power Virtual Server?*, found at:

  https://cloud.ibm.com/docs/power-iaas?topic=power-iaas-about-virtual-server
- IBM Cloud Pak® System Software documentation, found at:

  https://www.ibm.com/docs/en/cloud-pak-system-software

IBM offers cloud suites that help optimize business environments for reliability and better performance. IBM Cloud options are available to meet your expectations and needs based on the running infrastructure.

Figure 1-10 shows IBM different cloud options.

Figure 1-10 IBM Cloud options

For a private cloud, IBM brings the speed, agility, and pricing flexibility of public cloud solutions to on-premises. With the Power Private Cloud with Shared Utility Capacity offering, IBM enables by-the-minute resource sharing across systems and metering, which leads to unmatched flexibility, simplicity, and economic efficiency. The IBM Power Private Cloud Rack solution offers an open hybrid cloud offering based on IBM and Red Hat portfolios.

Figure 1-11 shows the IBM Power Private Cloud Rack solution.

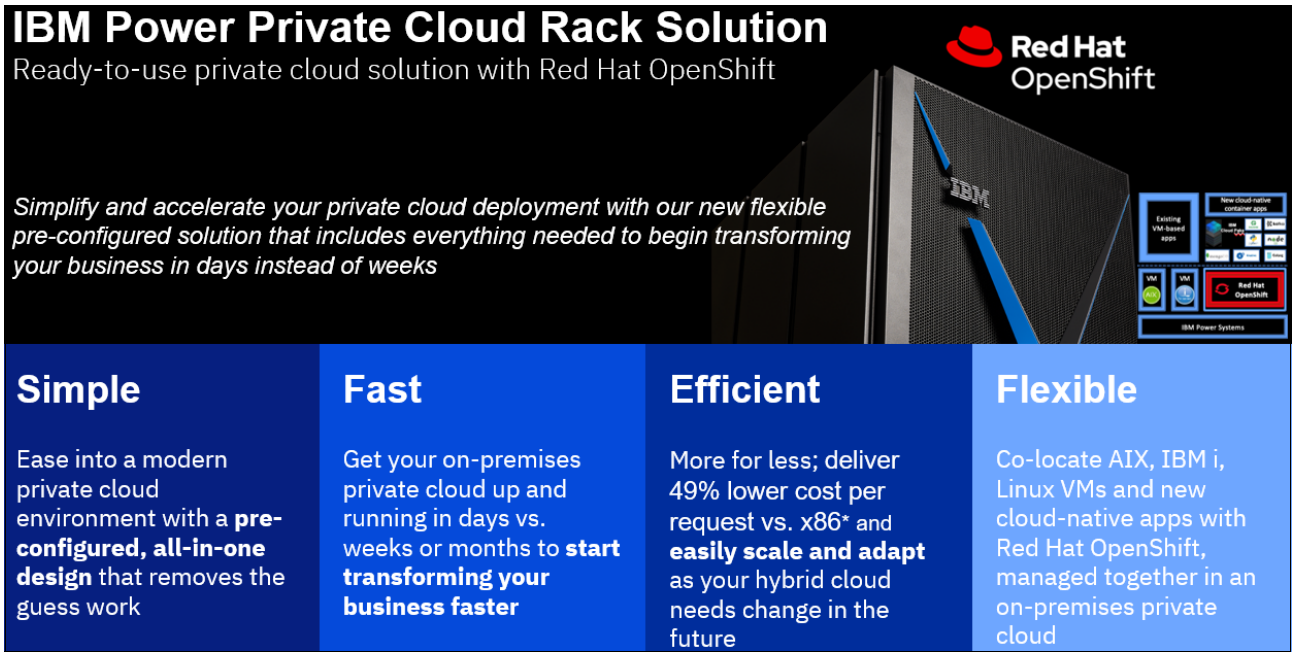


Figure 1-11 IBM Power Private Cloud Rack solution

When it comes to infrastructure as a service (IaaS) and private cloud management for Power servers, IBM offers IBM PowerVC with a fully automated platform.

Figure 1-12 shows IBM PowerVC for Power Private Cloud.

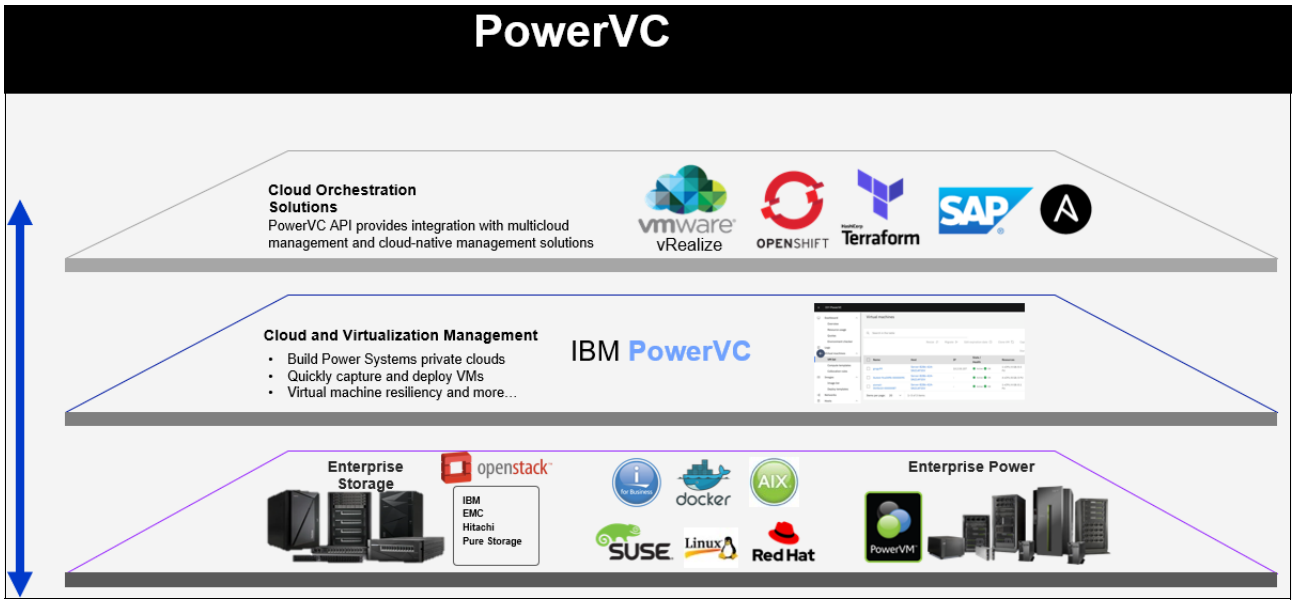


Figure 1-12 IBM PowerVC

If you seek an extensive and mixed cloud environment, hybrid cloud solutions can integrate private cloud services, on-premises, and public cloud services and infrastructure. They provide orchestration, management, and application portability across all layers. The result is a single, unified, and flexible distributed computing environment where an organization can run and scale its traditional or cloud-native workloads on the most appropriate computing model.

Figure 1-13 shows a high-level picture of hybrid cloud integration.

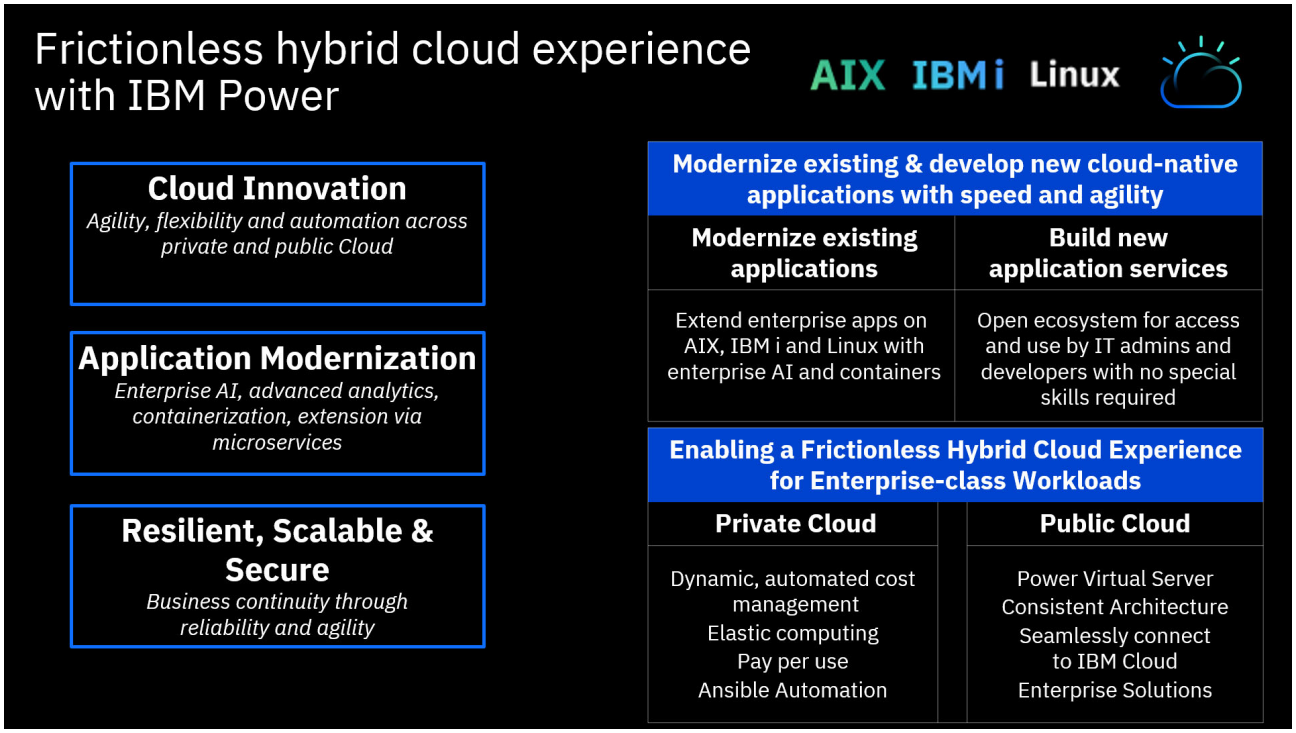


Figure 1-13 IBM hybrid cloud

1.8 PowerVC enhanced benefits

PowerVC is an advanced enterprise virtualization management offering for Power servers that is based on OpenStack technology. PowerVC is available to every PowerVM client. PowerVC provides simplified virtualization management and cloud deployments for AIX, IBM i, and Linux VMs that run on Power servers. Although PowerVM is virtualization for Power processor-based systems, PowerVC provides advanced virtualization and cloud management capabilities.

PowerVC is designed to build private cloud capabilities on Power servers and improve administration productivity. PowerVC provides fast VM image capture, deployment, resizing, and management. PowerVC includes integrated management of storage, network, and compute resources that simplifies administration. PowerVM with PowerVC also improve usage and reduce complexity.

PowerVC integrates with widely accepted cloud and automation management tools in the industry like Ansible, Terraform, and Red Hat OpenShift. It can be integrated into orchestration tools like IBM Cloud Automation Manager (CAM), VMware vRealize, or SAP Landscape Management (LaMa). One of the great benefits is the easy transition of VM images between private and public cloud, as shown in Figure 1-12 on page 28.

Several different options exist to set up communication between managed hosts and PowerVC, depending on whether the HMC or PowerVM NovaLink is used to communicate with hosts.

Attention: If PowerVM NovaLink is installed in the host system, the system still can be added for HMC management by normal procedure. However, if PowerVM NovaLink is installed in the host system and it is HMC-connected, the management type from PowerVC for this host always must be PowerVM NovaLink. HMC can be used to manage the hardware and firmware on the host system.

PowerVC is used widely by Power customers. PowerVC can manage up to 10,000 VMs; includes Multifactor Authentication (MFA) support; and supports persistent memory, creation of volume clones for backup, dynamic resource optimization, and single click server evacuation.

For more information, see IBM PowerVC, found at:

https://www.ibm.com/products/powervc

2

IBM PowerVM features in details

This chapter provides a detailed overview of the major PowerVM features. It is organized by grouping the capabilities into the following sections and describing a range of technologies that are associated with them:

- Processor virtualization
- Memory virtualization
- Storage virtualization
- Network virtualization

Then, this chapter describes further capabilities that can be used to provide greater flexibility, availability, and disaster recovery (DR) in the following sections:

- Dynamic logical partitioning
- Partition mobility
- Simplified Remote Restart
- VM Recovery Manager
- Capacity on Demand
- Power Enterprise Pools

2.1 Processor virtualization

Virtualization is a process that allows for more efficient utilization of physical computer hardware, and it is the foundation of cloud computing.

The main benefits of processor virtualization are as follows:

- Resource efficiency
- Simplified management
- Minimal downtime
- Faster provisioning

PowerVM is designed to use efficiently server resources when it pools resources and optimizes their usage across multiple application environments. This goal is achieved through processor virtualization, memory virtualization, and I/O virtualization.

Processor virtualization enables an operating system to use the CPU power more effectively and efficiently. It also helps to consolidate workloads into fewer systems, reduce administrative overhead, and better use hardware resources, all of which leads to lower costs. With PowerVM on IBM Power servers, you have the power and flexibility to address multiple system requirements in a single machine. IBM Micro-Partitioning supports multiple VMs per processor core. Depending on the Power server model, up to 1000 VMs can run on a single server, each with its own processor, memory, and I/O resources. Processor resources can be assigned at a granularity of 0.01 of a core.

Figure 2-1 reviews processor concepts.

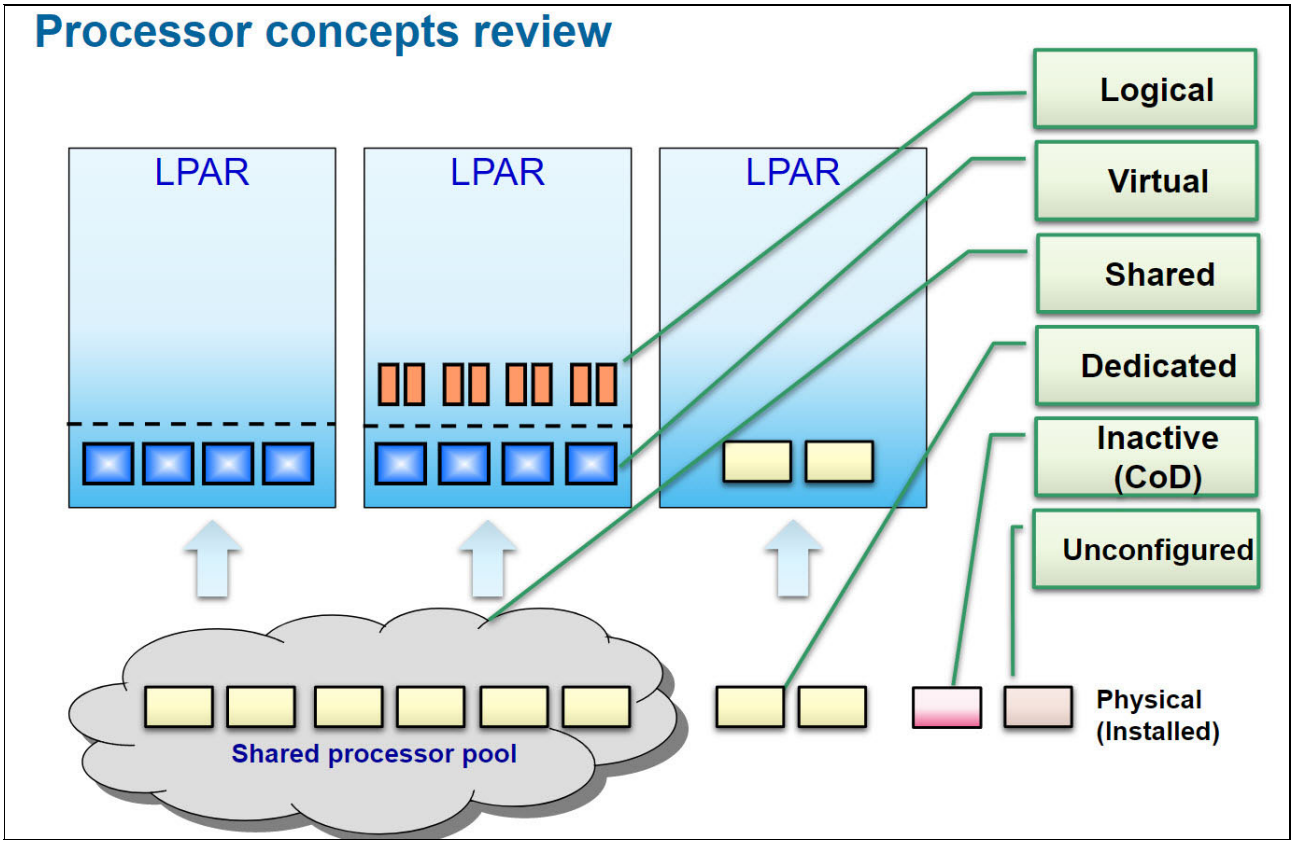


Figure 2-1 Processor concepts

This example shows 10 physical processors. From left to right, the figure shows six processors in the shared processor pool (SPP), two processors that are dedicated to a partition, one inactive Capacity on Demand (CoD) processor, and one processor that was unconfigured due to detected errors. Moving up in the figure, you can see two shared processor partitions, each with four virtual processors.

2.1.1 Dedicated processors

Dedicated processors are whole processors that are assigned to a single logical partition (LPAR). If you choose to assign dedicated processors to an LPAR, you must assign processors in whole numbers with at least one processor. Likewise, if you choose to remove processors from a dedicated LPAR, you must remove processors in whole numbers starting with one processor.

By default, the resources of a dedicated processor partition are shared with the default processor pool when the partition is inactive. Idle dedicated processors of the inactive partition are automatically added to the default SPP of the machine, which allows uncapped LPARs to access the idle processors, which belong to an inactive dedicated processor partition. When the dedicated processor partition is activated, it regains its processors. If wanted, it is possible to alter this behavior by changing the Processor Sharing setting from **Allow when partition is inactive** in the Advanced settings section in the **Processor** tab of the partition.

It also is possible to allow sharing of idle processing capacity of an active dedicated processor partition by changing the Processor Sharing mode to **Allow always** or **Allow when partition is active**. You can change the processor sharing mode of the dedicated processor LPAR at any time without having to shut down and restart the LPAR.

2.1.2 Dedicated donating

In the dedicated processing mode, physical processors are assigned as a whole to LPARs. During low workload, the unused processing resources remain idle, causing lower LPAR CPU utilization. *Dedicated donating* refers to the ability to donate idle or spare processors that are owned by a dedicated LPAR to an SPP. The LPAR maintains absolute priority over the processors that it owns. The donated processors are returned instantaneously to the dedicated-donating LPAR when it needs them back. Enabling this feature can help to increase global system utilization.

A dedicated processor LPAR donates idle cycles to shared pools, as shown in Figure 2-2.

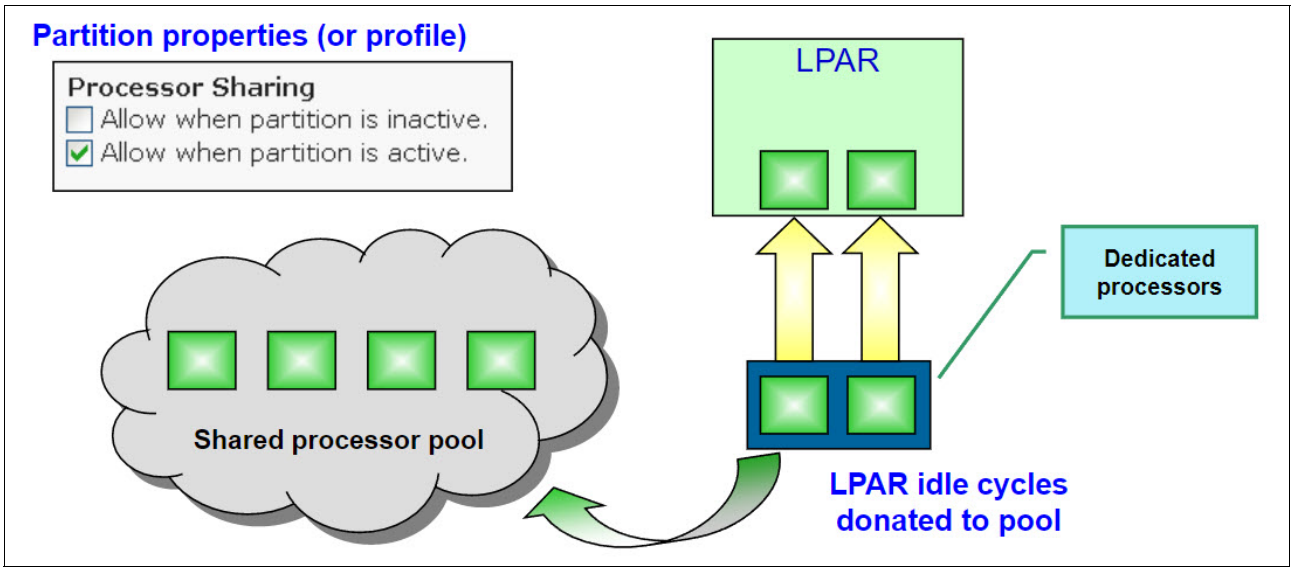


Figure 2-2 Donating mode

2.1.3 Shared processors

Shared processors are physical processors whose processing capacity is shared among multiple LPARs. The ability to divide physical processors and share them among multiple LPARs is known as Micro-Partitioning.

By default, all physical processors that are not dedicated to specific LPARs are grouped in an SPP. This pool is known as the default SPP, and it is automatically defined in the managed system. You can assign a specific amount of the processing capacity from the default SPP to LPARs that use shared processors. PowerVM also allows you to use the Hardware Management Console (HMC) to configure multiple shared processor pools (MSPP). These additional processor pools can be configured with a maximum processing unit value and a reserved processing unit value.

The maximum processing unit value limits the total number of processing units that can be used by the LPARs in the SPP. The reserved processing unit value is the number of processing units that are reserved for the usage of uncapped LPARs within the SPP.

You can assign partial processors to an LPAR that uses shared processors. Processing units are a unit of measure for shared processing power across one or more virtual processors. One shared processing unit on one virtual processor accomplishes approximately the same work as one dedicated processor.

As a best practice, the number of configured virtual processors must not exceed the maximum number so that server performance is not affected.

The maximum number of active virtual processors for a shared processor partition is limited by many factors. On IBM Power 870, IBM Power 880, IBM Power 870C, IBM Power 880C, IBM Power E980, and IBM Power E1080 model servers, the firmware has a limit of 128 active shared virtual processors per partition.

On all other models of IBM Power8, IBM Power9, and IBM Power10 processor-based servers, the firmware has a limit of 64 active shared virtual processors per partition.

Shared processor capacity is assigned in processing units from the shared processing pool:

- Minimum per partition is 0.05 processing units.
- More capacity is allocated in 0.01 processing unit increments.
- The partition-guaranteed amount is its entitled capacity.

Shared processors provide the following benefits:

- Configuration flexibility.
- Excess capacity might be used by other partitions.
- Time-sliced subprocessor allocations are dispatched according to demand and entitled capacity.
- A partition might run on multiple processors, based on interrupts and its entitled capacity.
- Reduce licensing cost and increase production throughput.

Partitions with shared processors are either capped or uncapped:

- A capped partition can use CPU cycles up to its entitled capacity. Excess cycles are ceded back to shared pool.
- Uncapped partitions share unused capacity based on a user-defined weighting. The weight scale is 0 - 255. If a partition needs extra CPU cycles, it can use unused capacity in the shared pool.

Figure 2-3 shows capped shared processors. A capped partition cannot use more than its entitled capacity regardless of how much idle capacity there is in the SPP. The figure illustrates a capped partition that uses all its entitled processor capacity twice over the time that is shown, but it cannot use more.

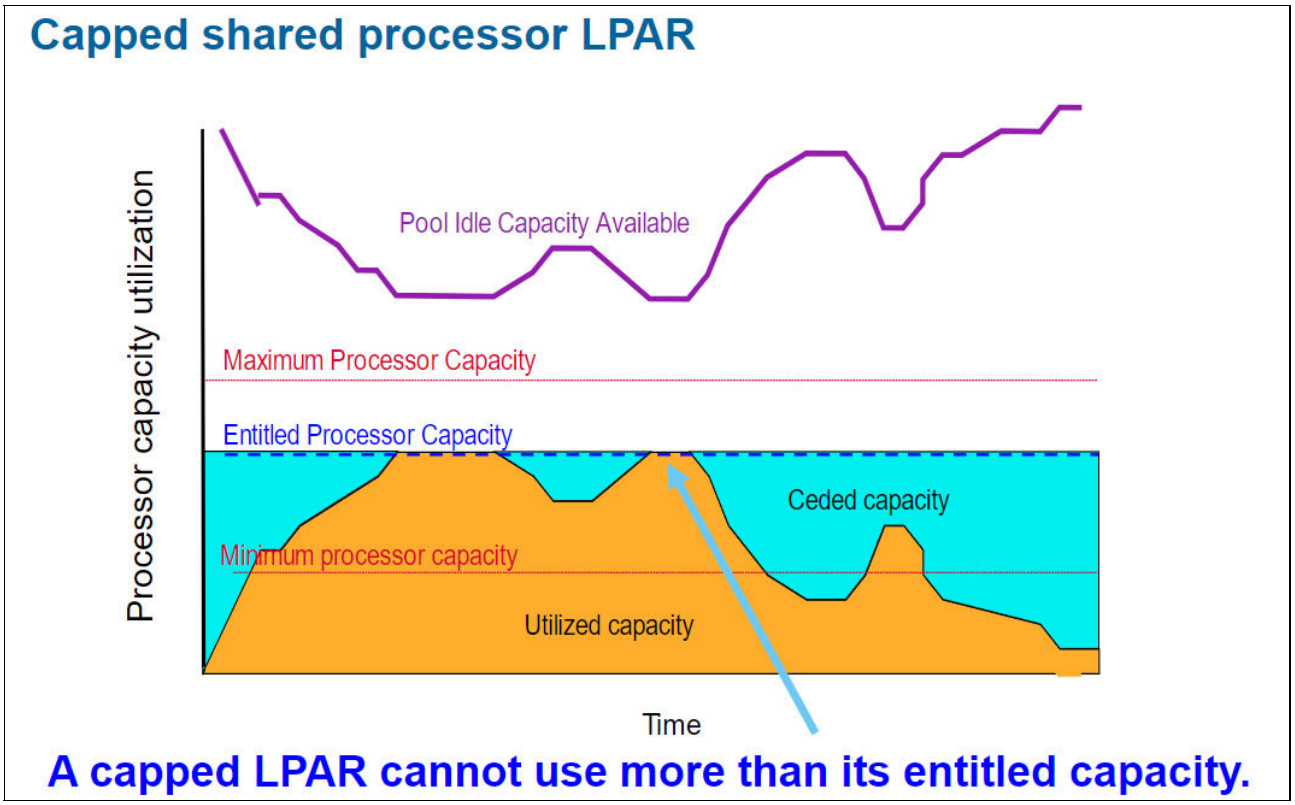


Figure 2-3 Capped shared processor LPAR

Figure 2-4 shows an uncapped partition that reaches its entitled capacity and is allowed to use excess capacity in the SPP. The partition can use more than its maximum processor capacity. The maximum setting limits dynamic LPAR operations only when changing the entitled capacity. It has no relevance to uncapped partitions that are allowed to use idle processing resources. Also, notice that the pool idle capacity is diminished because of the increased utilization by the uncapped partition. Uncapped partitions can grow only if there is excess capacity in the SPP and are eventually capped by their virtual processor setting.

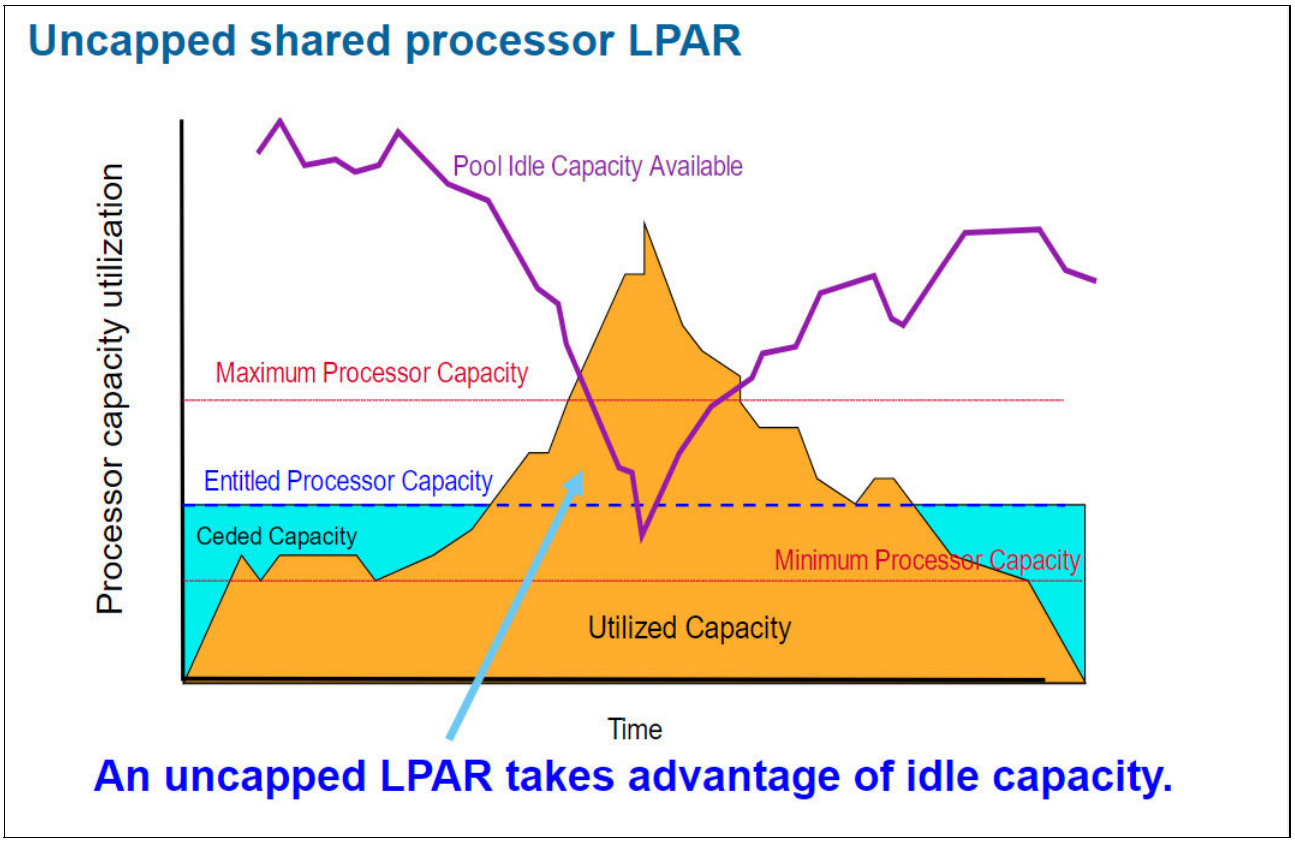


Figure 2-4 Uncapped shared processor LPAR

2.1.4 Virtual processors

A *virtual processor* is a representation of a physical processor core to the operating system of an LPAR that uses shared processors. It is the number of physical processors across which the LPAR can spread out. It represents the upper threshold for the number of physical processors that can be used. Each partition has its own assigned virtual processors. The partition works only on the virtual processors that are needed for the workload. Virtual processors of the partition, when not needed, are folded away by using the virtual processor folding feature.

2.1.5 Multiple shared processor pools

An SPP is a PowerVM technology that you can use to control the amount of processor capacity that partitions can use from the available physical processors in the system. This capability isolates workloads in an SPP and prevents the workload from exceeding an upper limit. This capability is also useful for software license management, where subcapacity licensing is involved.

Up to 64 SPPs can be defined on Power servers. A default SPP is automatically defined in the managed system. Each SPP has a maximum processing units value that is associated with it, as shown in Figure 2-5.

The maximum processing units define the upper boundary of the processor capacity that can be used by the set of partitions in the SPP. The system administrator optionally can allocate a number of reserved processing units to an SPP. The reserved processing units represent the available processor capacity beyond the processor capacity entitlements of the individual partitions in the SPP. The default value for the reserved processing units is zero.

By using the HMC, you can complete the following tasks:

- Allocate a specific amount of the processing capacity from the SPP to each partition that uses the shared processors.
- Configure the SPPs with a maximum processing unit value and a reserved processing unit value.
- View information about your SPP and change the properties of that pool.

Figure 2-5 provides an overview of MSPP.

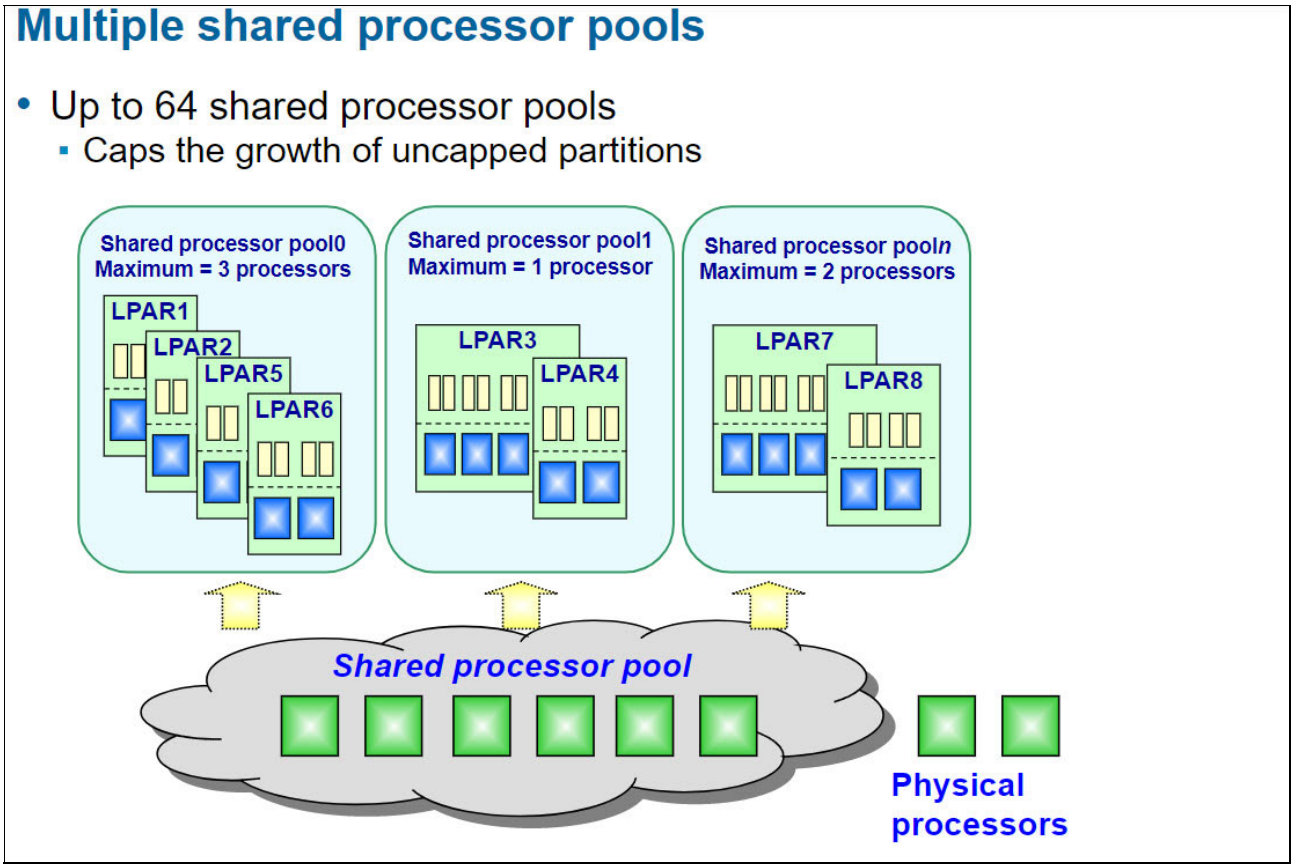


Figure 2-5 Multiple shared processor pools

SPPs are now available on IBM Power Systems Virtual Server (PowerVS).

Shared Processing Pools on PowerVS minimize software licensing costs and increase production throughput. Many software providers charge based on the number of available processors. SPPs enable users to combine applications together and define a maximum number of processing units for each pool.

For more information about the benefits of SPPs on PowerVS, see Reduce licensing costs and increase production throughput with Shared Processor Pools, found at:

https://www.ibm.com/cloud/blog/announcements/shared-processor-pools-on-ibm-power-systems-virtual-server

2.2 Memory virtualization

This section provides an overview of the memory virtualization capabilities like Active Memory Expansion (AME). It also provides an overview of how memory is addressed at the LPAR level.

2.2.1 Logical memory block

Memory is added and removed to and from LPARs in units of logical memory blocks (LMB). The memory block size can be changed by using the HMC enhanced GUI.

Each LPAR has a hardware page table (HPT). The HPT ratio is the ratio of the HPT size to the maximum memory value for the LPAR. The HPT is allocated in the server firmware memory overhead for the LPAR, and the size of the HPT can affect the performance of the LPAR.

Most servers are running with a 256 MB LMB size. Most systems default to this size, but customers can override the default. The LMB size is a tradeoff between granularity and internal management of the quantity of LMBs.

To select a reasonable logical block size for your system, consider both the performance that is wanted and the physical memory size. Use the following guidelines when you select logical block sizes:

- On systems with a small amount of memory that is installed (2 GB or less), a large LMB size results in the firmware taking an excessive amount of memory. Firmware must use at least one LMB. Generally, select the LMB size as no greater than one-eighth the size of the system's physical memory.
- On systems with a large amount of installed memory, small LMB sizes result in many LMBs. Because each LMB must be managed during boot, many LMBs can cause boot performance problems. Generally, limit the number of LMBs to 8 K or fewer.

A Power10 processor-based server does not support the smaller sizes, and it supports only 128 MB and 256 MB. If you use Live Partition Mobility (LPM) between older generations of Power servers and Power10 processor-based servers, ensure that you are using 128 MB or 256 MB to migrate to a Power10 processor-based server. To prepare for migration to a Power10 processor-based server, during a planned outage of the current server, change to a 256 MB LMB size if a smaller size is used. 256 MB allows partition migration between Power10, Power9, and Power8 processor-based servers.

For more information, see the following resources:

- Memory, found at:

  https://www.ibm.com/docs/en/power10?topic=resources-memory

- PowerVM features in the new Power10 servers, found at:

  https://community.ibm.com/community/user/power/blogs/pete-heyrman1/2021/09/27/powervm-features-in-the-new-power10-servers

- Changing the logical-memory block size, found at:

  https://www.ibm.com/docs/en/power10/9080-HEX?topic=options-changing-logical-memory-block-size

- *Power10 PowerVM Overview*, found at:

  https://public.dhe.ibm.com/systems/power/community/aix/PowerVM_webinars/110_P10_PowerVM.pdf

2.2.2 Active Memory Expansion

The AME feature is supported on Power8, Power9, and Power10 processor-based servers with AIX 7.1, AIX 7.2, and AIX 7.3.

With AME on Power servers, you can expand the amount of memory that is available to an AIX LPAR beyond the limits that are specified in the partition profile.

AME is an innovative technology that supports the AIX operating system. It helps enable the effective maximum memory capacity to be larger than the true physical memory maximum. Compression and decompression of memory content can enable memory expansion up to 100% or more. This expansion can enable a partition to complete more work or support more users with the same physical amount of memory. Similarly, it can enable a server to run more partitions and do more work for the same physical amount of memory.

AME uses CPU resources to compress and decompress the memory contents. The tradeoff of memory capacity for processor cycles can be an excellent choice, but the degree of expansion varies depending on how compressible the memory content is. The expansion also depends on having adequate spare CPU capacity that is available for this compression and decompression.

Servers like Power E1080 include a hardware accelerator that is designed to boost AME efficiency and use fewer processor core resources.

The AME feature can be enabled and disabled for a single AIX partition. On a Power server, a user can have a combination of LPARs, where AME is enabled for some and disabled for others.

Note: AME is not supported on IBM i and Linux operating systems.

Before you activate AME, ensure that the configuration requirements for AME are in place. As a best practice, use the AME planning tool, which helps plan the usage of AME for an existing workload.

For more information, see Preparing to configure Active Memory Expansion, found at:

https://www.ibm.com/docs/en/power10/9105-42A?topic=partitions-preparing-configure-active-memory-expansion

Key concepts and terminology

When a partition is configured with AME, the following two settings define how much memory is available:

- Physical memory

  The amount of physical memory that is available to the partition. Usually, it corresponds to the wanted memory in the partition profile.

- Memory expansion factor

  Defines how much of the physical memory is expanded.

Tip: The memory expansion factor can be defined individually for each partition.

The amount of memory that is available to the operating system can be calculated by multiplying the physical memory with the memory expansion factor. For example, in a partition that has 10 GB of physical memory and is configured with a memory expansion factor of 1.5, the operating system sees 15 GB of available memory.

Figure 2-6 shows an example of AME.

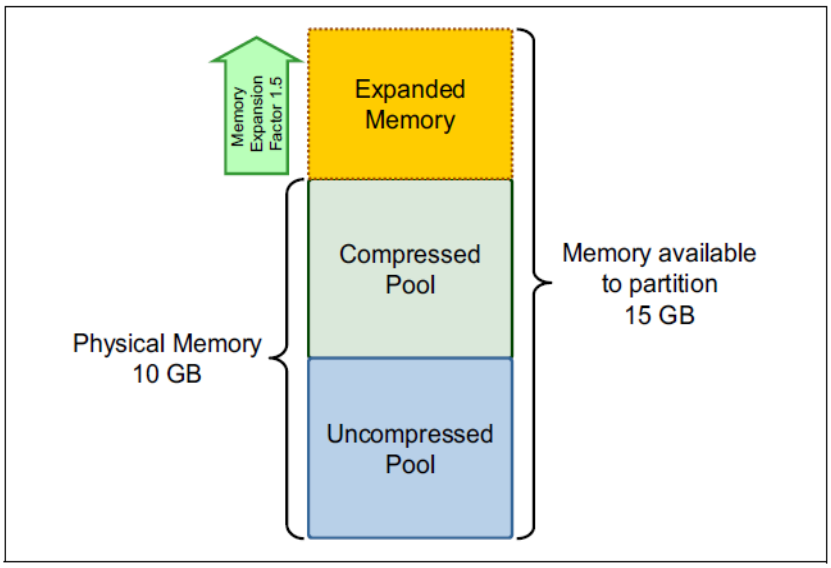


Figure 2-6 Active Memory Expansion example

The partition has 10 GB of physical memory that is assigned. It is configured with a memory expansion factor of 1.5. This configuration results in 15 GB of memory that is available to the operating system that runs in the partition. The physical memory is separated into the following two pools:

- Uncompressed pool

  Contains the noncompressed memory pages that are available to the operating system, much like normal physical memory.

- Compressed pool

  Contains the memory pages that were compressed by AME.

Some parts of the partition's memory are in the uncompressed pool, and others in the compressed pool. The size of the compressed pool changes dynamically.

Depending on the memory requirements of the application, memory is moved between the uncompressed and compressed pools.

When the uncompressed pool is full, AME compresses pages that are infrequently used and moves them to the compressed pool to free memory in the uncompressed pool.

When the application references a compressed page, AME decompresses it and moves it to the uncompressed pool. The pools, and the compression and decompression activities that take place when pages are moved between the two pools, are transparent to the application.

AME does not compress file cache pages and pinned memory pages.

If the expansion factor is too high, the target's expanded memory size cannot be achieved and a memory deficit forms. The effect of a memory deficit is the same as the effect of configuring a partition with too little memory. When a memory deficit occurs, the operating system might have to resort to paging out virtual memory to the paging space.

2.3 Storage virtualization

The following options are two of the most popular ways to provision storage to servers:

- Integrated disks.

  Integrated server disks are growing larger, which leads to requiring fewer disks for a specific amount of storage. A significant cost can be associated with the adapters and the attachment of these disks to servers. With such large disks, it is also more difficult to use all the available space.
- External storage subsystems, for example, storage area network (SAN) disks or network-attached storage (NAS) disks.

  The introduction of larger and cheaper disks drives down the costs per gigabyte of storage. The cost of adapters (and any switches and cabling) represents a significant investment if several servers are involved.

In many cases, it is beneficial to consolidate storage traffic through a single adapter to better use the available bandwidth. Cost savings include the server-related costs such as adapters or I/O slots, and the costs of the following items:

- Switches and switch ports (SAN or Etherchannel)
- Cables
- Installation of cables and patching panels

The cost benefits of storage virtualization can be quickly realized, and that is before any other benefits from the simplification of processes or organization in the business are considered.

A key aspect is how storage can be virtualized and mapped on client partitions. On PowerVM, the following features are available to map and virtualize storage:

- Virtual SCSI
- Virtual Fibre Channel
- Shared storage pools

2.3.1 Virtual SCSI

The Virtual I/O Server (VIOS) supports Virtual SCSI (vSCSI) disks, which are backed by various types of physical devices. Regardless of how the vSCSI disk is implemented by using various types of backing storage, all standard SCSI conventional rules apply to the device. The vSCSI disk behaves as a standard SCSI-compliant device. When a virtual disk is assigned to a client partition, the VIOS must be available before the client partitions can access it.

Regardless of the type of backing storage (physical volume, logical volume, or file) that is used for a vSCSI disk, the client partition recognizes it as a generic SCSI disk device. As such, a vSCSI disk can be used in the same way as a physical disk.

IBM PowerVM hypervisor (PHYP) provides the communications channels that permit client vSCSI adapters to communicate with server vSCSI adapters. Virtual target devices (VTDs) are used to allow backing storage that is physically owned by the VIOS to be presented as a target on a vSCSI bus. A client partition that uses a vSCSI bus can detect and use all virtual targets on that bus.

Figure 2-7 illustrates how multiple VTDs and their associated backing storage devices can be presented on a single vSCSI bus.

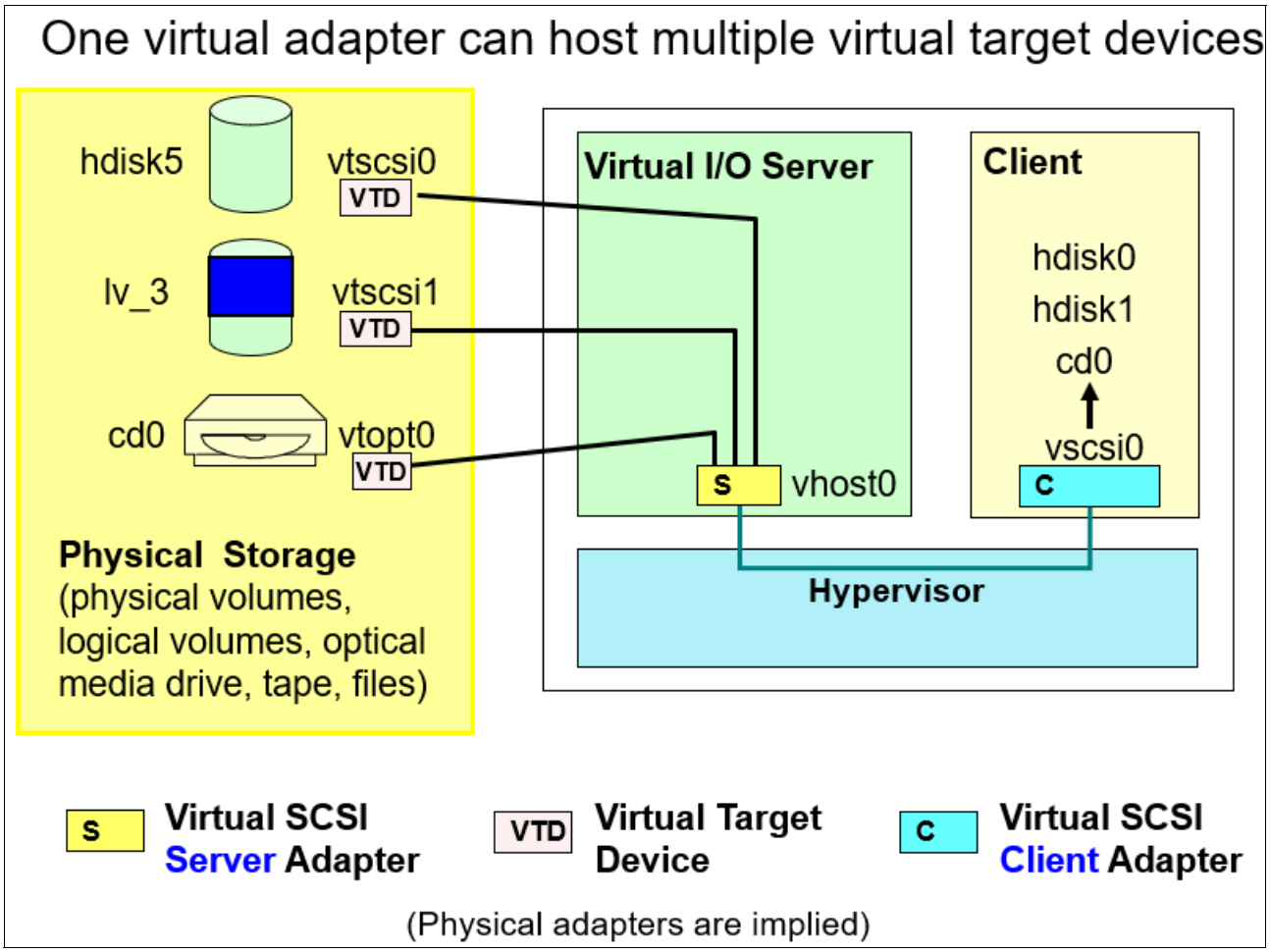


Figure 2-7 vSCSI overview

This approach can simplify configurations by reducing the number of virtual adapters. In the figure, the three backing storage devices map to the three virtual devices in the client LPAR. The devices that associate with the vSCSI server adapter (vhost#) can be changed dynamically. The client partition must run the **cfgmgr** command to see new devices.

Figure 2-8 shows two VIOS partitions that provide access to the same physical disk or LUN, for a single vSCSI client. In this case, it is necessary to use whole physical volumes (SAN LUNs) as backing devices for the virtual server SCSI adapters. It is not possible to use logical volumes or files.

In such a case, multipath input/output (MPIO) is used on the client partitions to provide increased resilience, serviceability, and scalability. It is not required to use MPIO on the VIOS partitions. Both MPIO configurations (client and VIOS) are shown in Figure 2-8.

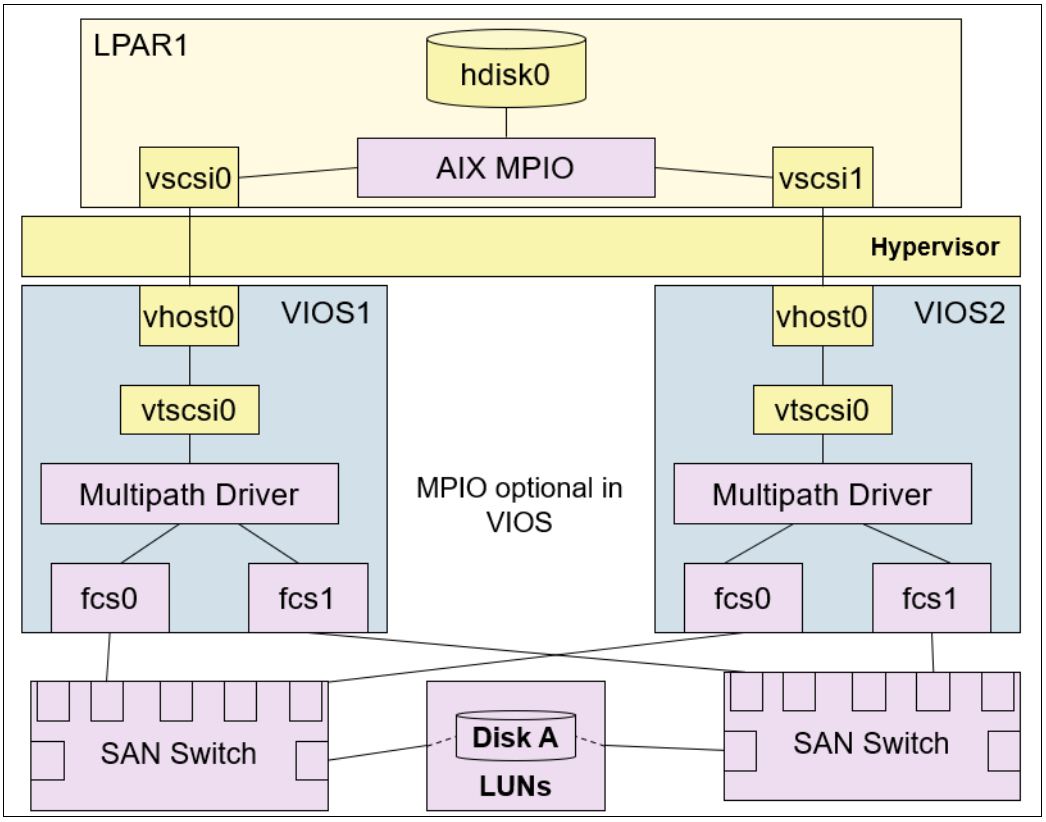


Figure 2-8 Dual Virtual I/O Servers with client MPIO

2.3.2 Virtual Fibre Channel

With N_Port ID Virtualization (NPIV), the role of the VIOS is fundamentally different because the VIOS facilitates adapter sharing only. No device-level abstraction or emulation exists. Rather than a storage virtualizer, the VIOS serving NPIV is a pass-through, which provides a Fibre Channel Protocol (FCP) connection from the client to the SAN.

If you use vSCSI disks, you must use the VIOS to create each vSCSI VTD and map each one to the vhost adapter for each client. With NPIV, the SAN can zone storage to the client LPAR's worldwide port name (WWPN). You do not need to create any vSCSI VTDs on the VIOS for the client to see the storage, which reduces the amount of VIOS management that is needed. Like vSCSI, you can configure the virtual Fibre Channel (VFC) adapters by using the HMC.

Using NPIV to access disks provides several benefits over vSCSI. Some benefits are related to increased functions, such as the ability for the client operating system to see the real type of disk that is being used. Client-level tools often need this capability for certain functions to work.

Some benefits are related to performance, such as the ability to load more sophisticated Performance and Capacity Monitors (PCMs) in the client, which provide load-balancing or other benefits when MPIO is used.

Some benefits apply to ease of administration, such as removing the need to document all the client endpoint devices on the VIOS to prevent accidental overwriting of data. With LPM, any vSCSI disks that are used by a client partition must be accessible by all VIOS partitions that are involved in a relocation.

NPIV provides redundancy for all components except for the disk itself. Redundancy for the disk can be provided by the SAN storage or implemented by using disk mirroring along with the MPIO configuration.

Figure 2-9 represents the minimum MPIO configuration with two paths to the same disk over two VFC client adapters that are connected through physical FC adapters on the dual VIOS.

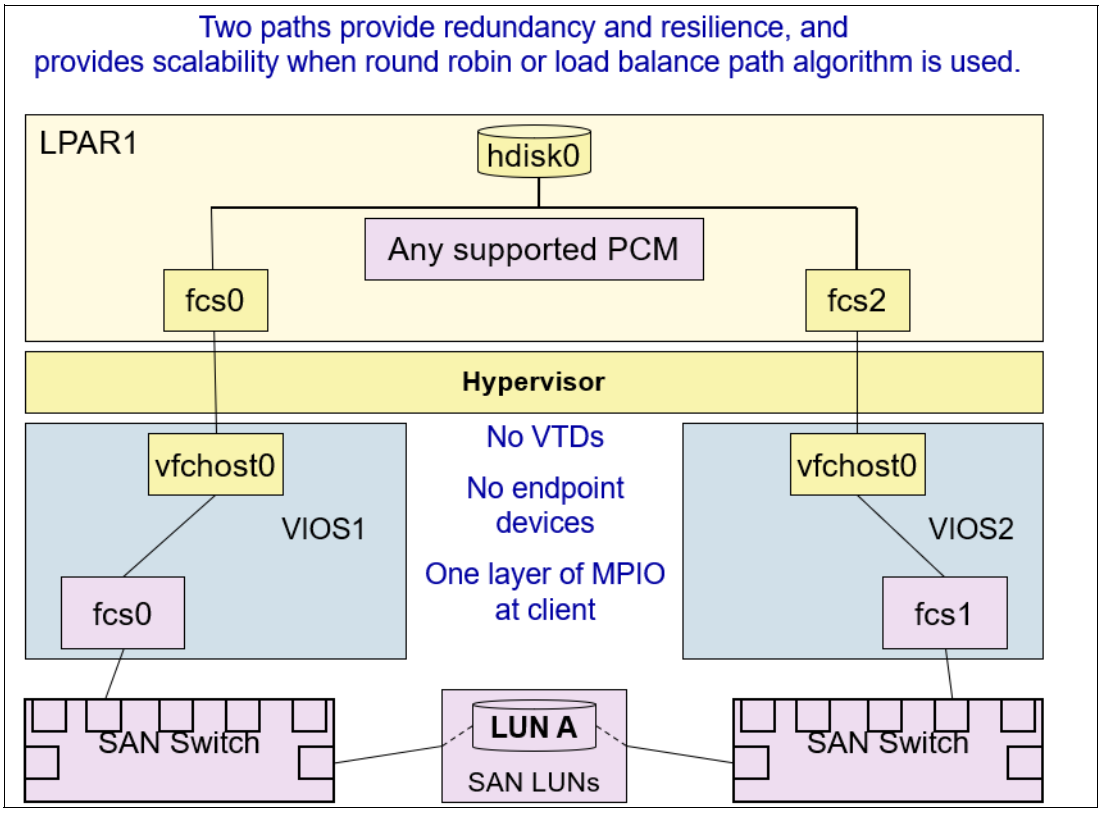


Figure 2-9 Client with two paths to the same disk

2.3.3 Shared storage pools

A *VIOS storage group*, also known as a VIOS cluster, consists of 1 - 24 VIOS partitions, which share a single storage pool. Storage is provisioned from the shared storage pool (SSP) for clients and uses the vSCSI protocol. An SSP is a pool of SAN storage devices that can be used among VIOSs. It is based on a cluster of VIOSs and a distributed data object repository with a global namespace. Each VIOS that is part of a cluster represents a cluster node.

The VIOS storage group can be managed from the HMC Enhanced+ GUI, the VIOS command-line interface (CLI), or from the **cfgassist** utility. The HMC also can be used to create backing storage and map it to clients. VIOS nodes in a storage group can continue to provide traditional virtual services such as vSCSI, VFC, and Shared Ethernet Adapter (SEA) services that are separate from the cluster.

Figure 2-10 shows VIOS cluster nodes that have a global view into the SSP.

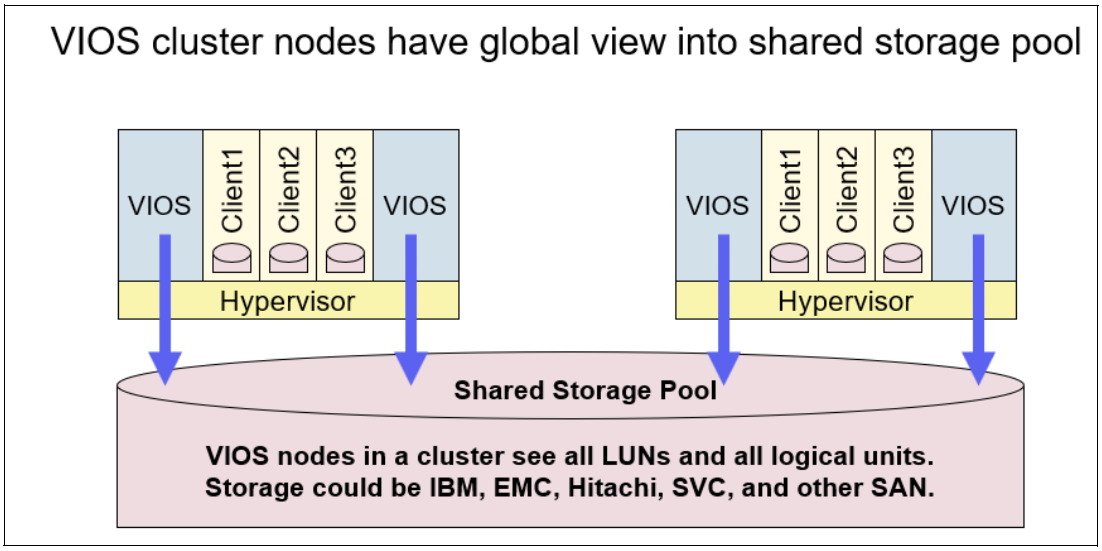


Figure 2-10 VIOS shared storage pool

For more information, see Shared storage pool clusters, found at:

https://www.ibm.com/docs/en/power10/9043-MRX?topic=storage-shared-pool-clusters

2.4 Network virtualization

Virtual networks bring many benefits to modern enterprises. Network virtualization capabilities are of paramount importance when you consider the network strategy for an enterprise.

The main benefits from virtual networks are as follows:

- Consolidation of separate networks
- Regulatory compliance
- Infrastructure integration for mergers and acquisitions
- Isolation of critical resources
- Cost savings

Software-defined networking (SDN) is a type of network virtualization that virtualizes hardware that controls network traffic routing (called the “control plane”). Network function virtualization (NFV) virtualizes one or more hardware appliances that provide a specific network function (for example, a firewall, load balancer, or traffic analyzer), which makes those appliances easier to configure, provision, and manage.

2.4.1 Virtual Ethernet Adapter

With virtual Ethernet, LPARs can communicate with each other without having to assign physical hardware to the LPARs.

Virtual Ethernet Adapters (VEAs) can be created on LPARs and connected to virtual LANs. TCP/IP communications over these virtual LANs are routed through the server firmware.

VEAs are connected to an IEEE 802.1q (virtual local area network (VLAN))-style virtual Ethernet switch. By using this switch function, LPARs can communicate with each other by using VEAs and assigning VLAN IDs that enable them to share a common logical network. The VEAs are created and the VLAN ID assignments are done with the HMC. The system transmits packets by copying the packet directly from the memory of the sender LPAR to the receive buffers of the receiver LPAR without any intermediate buffering of the packet.

You can configure an Ethernet bridge between the virtual LAN and a physical Ethernet adapter that is owned by a VIOS or IBM i LPAR. The LPARs on the virtual LAN can communicate with an external Ethernet network through the Ethernet bridge. You can reduce the number of physical Ethernet adapters that are required for a managed system by routing external communications through the Ethernet bridge. VEAs are connected to the hypervisor Ethernet switch, as shown in Figure 2-11.

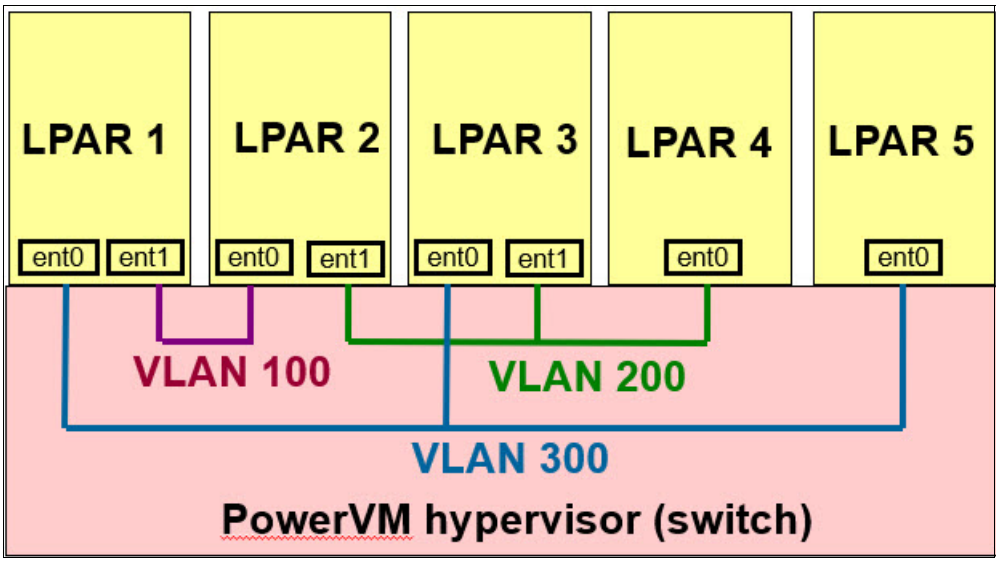


Figure 2-11 Virtual Ethernet switch

The number of VEAs that are allowed for each LPAR varies by operating system. AIX supports up to 256 VEAs for each LPAR. Version 2.6 of the Linux kernel supports up to 32,768 VEAs for each LPAR. Each Linux LPAR can belong to a maximum of 4,094 virtual LANs.

Apart from a Port VLAN ID, the number of extra VLAN ID values that can be assigned for each VEA is 19. This value indicates that each VEA can be used to access 20 networks. The HMC generates a locally administered Ethernet MAC address for the VEAs so that these addresses do not conflict with physical Ethernet adapter MAC addresses.

After a specific virtual Ethernet is enabled for an LPAR, a network device is created in the LPAR. This network device is named ent*X* on AIX LPARs, CMN*XX* on IBM i LPARs, and eth*X* on Linux LPARs, where *X* represents sequentially assigned numbers. The user can set up TCP/IP configuration like a physical Ethernet device to communicate with other LPARs.

2.4.2 Shared Ethernet Adapter

A *SEA* is a VIOS component that bridges a physical Ethernet adapter and one or more VEAs. SEAs on the VIOS LPAR allow VEAs on client LPARs to send and receive outside network traffic.

The real adapter can be a physical Ethernet adapter, a link aggregation (LA) or Etherchannel device, or a single-root I/O virtualization (SR-IOV) logical port (LP). The real adapter cannot be another SEA, or a VLAN pseudo-device. The VEA must be a virtual I/O Ethernet adapter. It cannot be any other type of device or adapter.

The SEA enables LPARs on the virtual network to share access to the physical network and communicate with stand-alone servers and LPARs on other systems. It eliminates the need for each client LPAR to own a real adapter to connect to the external network.

A SEA provides access by connecting the internal VLANs with the VLANs on the external switches. This approach enables LPARs to share the IP subnet with stand-alone systems and other external LPARs. The SEA forwards outbound packets that are received from a VEA to the external network. It forwards inbound packets to the appropriate client LPAR over the virtual Ethernet link to that partition. The SEA processes packets at layer 2, as shown in Figure 2-12. The original MAC address and VLAN tags of the packet are visible to other systems on the physical network.

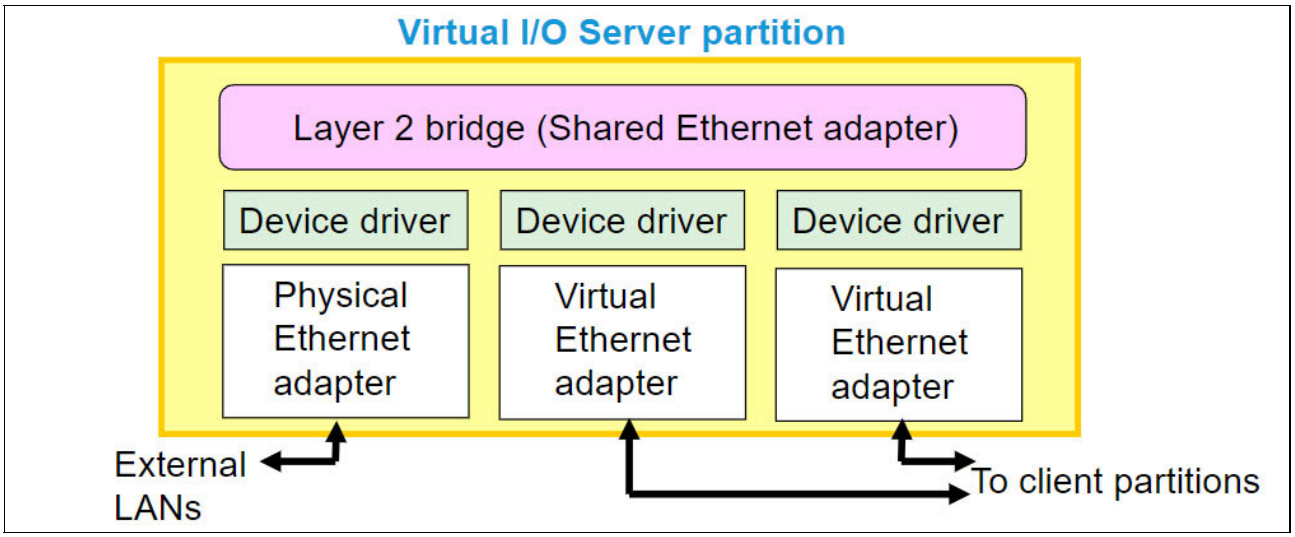


Figure 2-12 SEA layer 2 bridge

2.4.3 Single-root I/O virtualization

SR-IOV is a Peripheral Component Interconnect Express (PCIe) standard architecture that defines extensions to PCIe specifications. It enables multiple LPARs with multiple operating systems that run simultaneously within a system to share PCIe devices.

SR-IOV enables PCIe adapters to become self-virtualizing. It enables adapter consolidation, through sharing, much like logical partitioning enables server consolidation. SR-IOV is analogous to Micro-Partitioning. It is like a hardware “slice” of an adapter down to the wire. With an adapter capable of SR-IOV, you can assign virtual slices of a single physical adapter to multiple partitions through LPs. SR-IOV does not require a VIOS. You can use SR-IOV along with VIOS on the same system.

Overall, SR-IOV provides integrated virtualization without VIOS and with greater server efficiency as more of the virtualization work is done in the hardware and less in the software.

Several partitions can share an SR-IOV adapter for Ethernet connections without the usage of the VIOS. The number of partitions is adapter-dependent. Partition workloads also can impact the number of partitions that an adapter and physical port support.

SR-IOV requires specific hardware, software, and firmware levels. An Ethernet network interface card (NIC) can be used with SR-IOV.

SR-IOV-capable adapters support SR-IOV shared mode or dedicated mode:

- If the adapter is in *dedicated mode*, the adapter is either unowned or is owned by a single LPAR. When an LPAR owns an adapter in dedicated mode, it owns all physical ports on the adapter, and it is managed by the OS of the partition.
- In *shared mode*, the adapter is owned by the hypervisor. LPs can be created and assigned to partitions. These LPs allow partitions to access a share of a particular port on the adapter. In shared mode, virtual Network Interface Controllers (vNICs) can be created to support LPM and Simplified Remote Restart (SRR) capabilities in the SR-IOV client.

In Figure 2-13, each LPAR shares the bandwidth of a physical port on two different physical SR-IOV adapters. The percentage of the port that is shared is determined at the configuration of the LP, but the percentage cannot exceed an aggregate usage of 100%.

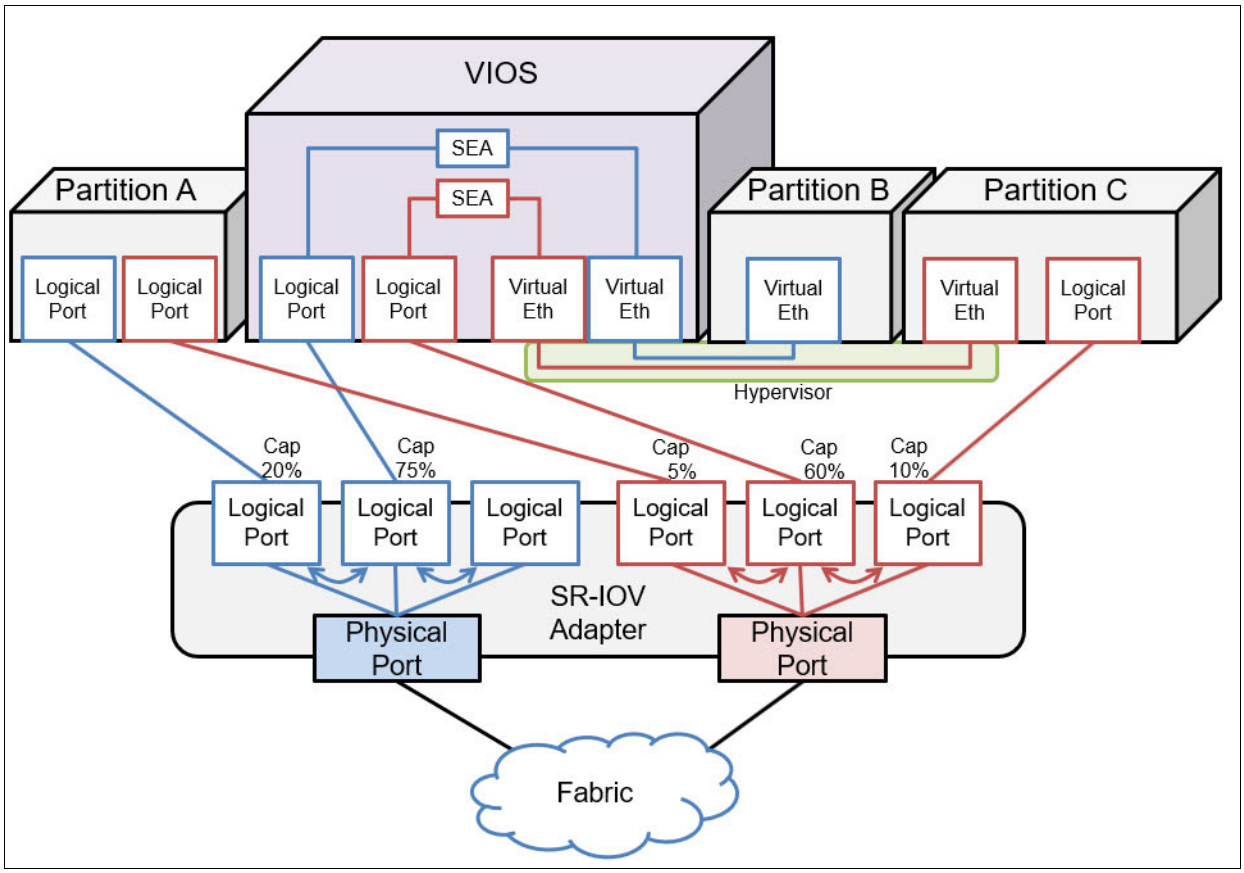


Figure 2-13 SEA with multiple LPARs

The main benefits of SR-IOV are as follows:

- Direct-access I/O and performance. The primary benefit of allocating adapter functions directly to a partition, as opposed to using virtualization like VIOS, is performance. The processing overhead that is involved in passing client instructions through a virtual intermediary (VI) to the adapter and back is substantial. With direct-access I/O, SR-IOV-capable adapters that run in shared mode allow the operating system to directly access the slice of the adapter that is assigned to its partition. There is no control or data flow through the hypervisor.
- From the partition perspective, the adapter appears to be physical I/O. Regarding CPU and latency, the adapter exhibits characteristics of physical I/O. Because the operating system is directly accessing the adapter, if the adapter has special features, like multiple queue support or receive side scaling (RSS), the partition also can use them if the operating system includes those capabilities in its device drivers.
- Adapter sharing. The current trend of consolidating servers to reduce cost and improve efficiency increases the number of partitions per system, which drives a requirement for more I/O adapters per system to accommodate them. SR-IOV addresses and simplifies that requirement by enabling the sharing of SR-IOV-capable adapters. Because each adapter can be shared and directly accessed by several partitions, depending on the adapter, the partition to PCI slot ratio can be improved without adding the overhead of a VI. The number of partitions depends on the adapter. Partition workloads also can impact the number of partitions that an adapter and physical port support.
- Flexible deployment. A Power server's SR-IOV capability enables flexible deployment configurations, ranging from a simple single-partition deployment to a complex, multi-partition deployment that involves VIOS partitions and VIOS clients that run different operating systems.

  In a single-partition deployment, the SR-IOV-capable adapter in shared mode is wholly owned by a single partition, and no adapter sharing takes place. This scenario offers no practical benefit over traditional I/O adapter configuration, but the option is available. In a more complex deployment scenario, an SR-IOV-capable adapter might be shared by both VIOS and non-VIOS partitions. The VIOS partitions might further virtualize the LPs as SEAs for VIOS client partitions. This scenario leverages the benefits of direct-access I/O, adapter sharing, and quality of service (QoS) that SR-IOV provides. Other benefits include higher-level virtualization functions, such as LPM (for the VIOS clients) that VIOS can offer.
- Reduced cost. SR-IOV facilitates server consolidation by reducing the number of physical adapters, cables, switch ports, and I/O slots that are required per system. This reduction results in reduced cost in terms of physical hardware that is required, and reduced associated energy costs for power consumption, cooling, and floor space. You might save the extra cost on CPU and memory resources, compared to a VIOS adapter sharing solution. SR-IOV does not have the resource overhead that is inherent in using a virtualization intermediary to interface with the adapters.

Note: For more information about the adapters that are supported in SR-IOV shared mode for Power8, Power9, and Power10 processor-based servers, see PowerVM SR-IOV FAQs, found at:

https://community.ibm.com/HigherLogic/System/DownloadDocumentFile.ashx?DocumentFileKey=0f22677f-0a29-de26-171e-88318a77991e&forceDialog=0

2.4.4 SR-IOV with virtual Network Interface Controller

vNIC is a PowerVM virtual networking technology that delivers enterprise capabilities and simplifies network management. It is a high-performance, efficient technology that, when combined with SR-IOV, provides bandwidth control QoS capabilities at the virtual NIC level. vNIC reduces virtualization overhead, which results in lower latencies and fewer server resources (CPU and memory) that are required for network virtualization.

In addition to the improved virtual networking performance, the client vNIC can take full advantage of the QoS capability of the SR-IOV adapters that are supported on Power servers. Essentially, the QoS feature ensures that an LP receives its share of adapter resources, which includes its share of the physical port bandwidth.

vNIC is a type of VEA that is configured on the LPAR. Each vNIC is backed by an SR-IOV LP that is available on the VIOS. The key advantage of placing the SR-IOV LP on the VIOS is that it makes the client LPAR eligible for LPM.

SEA can also use the SR-IOV LP as its physical network device, but the LP must be configured with promiscuous mode enabled.

A vNIC that is combined with SR-IOV adapters provides the best of both QoS and flexibility.

The vNIC configuration requires the following firmware and operating system support:

- System firmware level FW840 and HMC 840 or later
- VIOS 2.2.4.0 or later
- vNIC driver support from AIX and IBM i systems

Note: SR-IOV adapters do not require VIOS. However, to configure vNIC, VIOS is required.

During LPM operations, HMC handles the creation of the vNIC server and backing devices on the target system. HMC also handles the cleanup of devices on the source system when LPM completes successfully. HMC has built-in capability to provide auto-mapping of backing devices and hosting VIOSs between the source and target servers.

The SR-IOV port label, the available capacity and the virtual functions (VFs) count, and the adapter and VIOS redundancy are some of the key factors that are used by the HMC for auto-mapping. Optionally, you can also specify your own mapping settings.

For more information, see the following resources:

- vNIC - Introducing a New PowerVM Virtual Networking Technology, found at:

  https://community.ibm.com/community/user/power/blogs/charles-graham1/2020/06/19/vnic-introducing-a-new-powervm-virtual-networking

- PowerVM vNIC and vNIC Failover FAQ, found at:

  https://community.ibm.com/HigherLogic/System/DownloadDocumentFile.ashx?DocumentFileKey=96088528-4283-8b61-38b0-a39c9ed990c7&forceDialog=0%3E%3E

2.4.5 Hybrid Network Virtualization

Hybrid Network Virtualization (HNV) is a new technology that is part of the PowerVM capabilities. IBM introduced production-ready HNV support with Power server firmware FW940.10 and HMC9.1.941.0, and the supporting AIX and IBM i releases.

IBM introduced Linux HNV support, with Power server firmware FW950.00 and HMC 9.2.950.0, along with supporting Linux releases.

HNV allows AIX, IBM i, and Linux partitions to leverage the efficiency and performance benefits of SR-IOV LPs and participate in mobility operations, such as active and inactive LPM and SRR. HNV is enabled by selecting a new **Migratable** option when an SR-IOV LP is configured.

HNV leverages existing technologies such as AIX Network Interface Backup (NIB), IBM i Virtual IP Address (VIPA), or Linux active-backup bonding. HNV introduces new automation for configuration and mobility operations to provide an integrated high-performance network adapter sharing capability with partition mobility.

The approach is to create an active-backup configuration within a partition where the primary device is an SR-IOV LP, and the backup device is a virtual device such as a VEA or vNIC. As the primary device, the SR-IOV LP provides high-performance, low-overhead network connectivity.

During an LPM operation, or when the primary device cannot provide network connectivity, network traffic flows through the backup virtual device.

If the partition is configured with a migratable SR-IOV LP, the HMC dynamically removes the SR-IOV LP as part of the migration operation. Then, network traffic is forced to flow through the virtual backup device. With the SR-IOV LPs removed, the HMC can migrate the partition. Before migration, the HMC provisions the SR-IOV LPs on the destination system to replace the previously removed LPs. When the partition is on the destination system, the HMC dynamically adds the provisioned LPs to the partition where they are integrated into the active-backup configuration.

The configuration of HNV devices is a two-step process:

1. The first step is to configure a migratable LP and its backup device at the HMC for a partition profile or a partition. The new devices become visible to the partition on activation of the partition profile or when the migratable LP is dynamically added to the partition.
2. When the devices are visible to the partition, OS-specific configuration steps are required to complete the configuration.

The minimum requirements for HNV are as follows:

- Power9 and Power10 processor-based servers.
- System firmware: FW940.10 or later. For Linux, FW950.00 or later.
- HMC version and release: HMC 9.1.941.0 or later. For Linux, HMC 9.2.950.0 or later.
- AIX 7.2 with the 7200-04 Technology Level and Service Pack (SP) 7200-04-02-2015.
- IBM i 7.3 TR7 and IBM i 7.4 TR1.
- VIOS 3.1.1.20.
- Red Hat Enterprise Linux 8.4.
- SUSE Linux Enterprise Server 15 SP 3 or later. Requires the NetworkManager package and NetworkManager service to be enabled.

- Linux support requirements:
  - Backup device that is limited to virtual Ethernet.
  - Supported adapter FCs EC2R/EC2S, EC2T/EC2U, EC3L/EC3M, or EC66/EC67.

For more information, see the following resources:

- HNV, found at:

  https://www.ibm.com/docs/en/linux-on-systems?topic=servers-hybrid-network-virtualization
- Hybrid Network Virtualization - Using SR-IOV for Optimal Performance and Mobility, found at:

  https://community.ibm.com/community/user/power/blogs/charles-graham1/2020/06/19/hybrid-network-virtualization-using-sr-iov-for-opt

2.5 Dynamic logical partitioning

Dynamic logical partitioning (DLPAR) has been one of the important features of Power servers and part of the PowerVM offering for many years. DLPAR is the capability of LPARs to move, add, or remove hardware resources on Power servers without shutting down the operating systems that run on the LPAR. The hardware resources can be processors, memory, I/O adapters, and virtual adapters. These resources are adjusted based on the demands of the LPAR.

You can perform the following basic operations with DLPAR:

- Move a resource between partitions on the same Power server.
- Remove resources from a partition.
- Add a resource to a partition.

Processors, memory, and I/O adapters that are not assigned to a partition appear as unassigned resources and can be viewed through HMC. A partition on the system has no visibility to the other partitions on the same system and unassigned resources.

Note: Virtual adapters can be added or removed only from the partitions, and they cannot be moved between partitions because they are virtual resources.

When you remove a processor from an active partition, the system releases it to the pool, and then that processor can be added to any active partition. When a processor is added to an active partition, it has full access to all the partition's memory, I/O address space, and I/O interrupts. The processor can participate in that partition's workload.

You can add or remove memory in multiple LMBs. When memory from a partition is removed, the time that it takes to complete a DLPAR operation is relative to the number of memory chunks that are removed.

The effects of memory removal on an application that runs on an AIX partition are minimized by the fact that the AIX kernel runs almost entirely in virtual mode. The applications, kernel extensions, and most of the kernel uses only virtual memory. When memory is removed, the partition might start paging because parts of the AIX kernel are pageable, which might degrade performance. When you remove memory, you must monitor the paging statistics to ensure that paging is not induced.

Note: Memory is managed by the hypervisor in blocks that are called *LMBs*. By default, the LMB size is 256 MB, but might be set to a smaller value depending on the model.

You can also add, move, or remove I/O adapters and virtual adapters, such as network adapters, FC adapters, tape drives, optical devices, virtual network adapters, virtual FC adapters, vSCSI, SR-IOV LPs, and vNIC adapters from active or inactive partitions.

Note: The physical hardware resources that are assigned as *required* in the LPAR profiles cannot be moved or removed until the partition profile is modified and saved.

DLPAR operations are performed by using the HMC GUI or CLI. DLPAR does not compromise the security of a partition. Resources that are moved in between partitions are reinitialized so that no residual data is left behind.

The DLPAR operations can be performed only when the Resource Monitoring and Control (RMC) state is active, which can be viewed from the HMC GUI or CLI. If the state is not active or inactive, DLPAR operations cannot be performed. RMC can be inactive due to several reasons. These reasons can be the RMC TCPIP communication between the HMC and LPAR, RMC port 657 is blocked on firewall, or Reliable Scalable Cluster Technology (RSCT) daemons are not running on partitions or are not synced with HMC. For more information about how to fix RMC errors, see Fixing the No RMC Connection Error, found at:

`https://www.ibm.com/support/pages/fixing-no-rmc-connection-error`

Resource Monitoring and Control

RMC is used by a management console to perform dynamic operations on an LPAR.

The RMC subsystem that is running on the HMC and the LPARs that are running AIX or Linux operating system or VIOS are responsible for establishing a management domain between the HMC and the LPARs that have the HMC as the Management Control Point (MCP). This management domain formation also can be referred as an RMC connection. It is the scalable, reliable backbone of RSCT.

The RMC subsystem facilitates the following functions:

- DLPAR operations.
- LPM, hibernation, and remote start.
- VIOS management operations.
- Operating system shutdown operation.
- CoD.
- Sending serviceable events from the operating system to the HMC. For AIX and Linux partitions, the serviceable events can be reported to your service provider.
- Partition inventory.
- AIX performance manager data collection.
- Code update.

An RMC connection also is used between the HMC and each LPAR that is running the IBM i operating system to facilitate the following functions:

- Sending serviceable events from the operating system to the HMC. For IBM i partitions, these serviceable events are informational for the HMC. Reporting the event to the service provider is the responsibility of current versions of IBM i.
- Partition inventory.
- Connection monitoring.

2.6 Partition mobility

Partition mobility is the ability to migrate AIX, IBM i, and Linux LPARs from one system to another one. The mobility process transfers the system environment, which includes the processor state, memory, attached virtual devices, and connected users.

Partition mobility and *partition migration* refer to the same capability.

By using *active partition migration* or *LPM*, you can migrate AIX, IBM i, and Linux LPARs that are running, including the operating system and applications, from one system to another one. The LPAR and the applications that are running on that migrated LPAR do not need to be shut down.

By using *inactive partition migration*, you can migrate a powered-off AIX, IBM i, or Linux LPAR from one system to another one.

Note: Since its inception, LPM has been accepted as the *de facto* terminology for all mobility operations, so it is used interchangeably for both active and inactive partition migration operations.

2.6.1 Live Partition Mobility

LPM is a PowerVM capability that allows you to move a running LPAR, including its operating system and running applications, from one system to another one without any shutdown or without disrupting the operation of that LPAR.

You can use the HMC to migrate an active or inactive LPAR from one server to another one.

Figure 2-14 on page 55 shows an example of LPM that is initiated from the HMC. You can move the client partition from the source server to the target server without disrupting the operating system and applications on the partition.

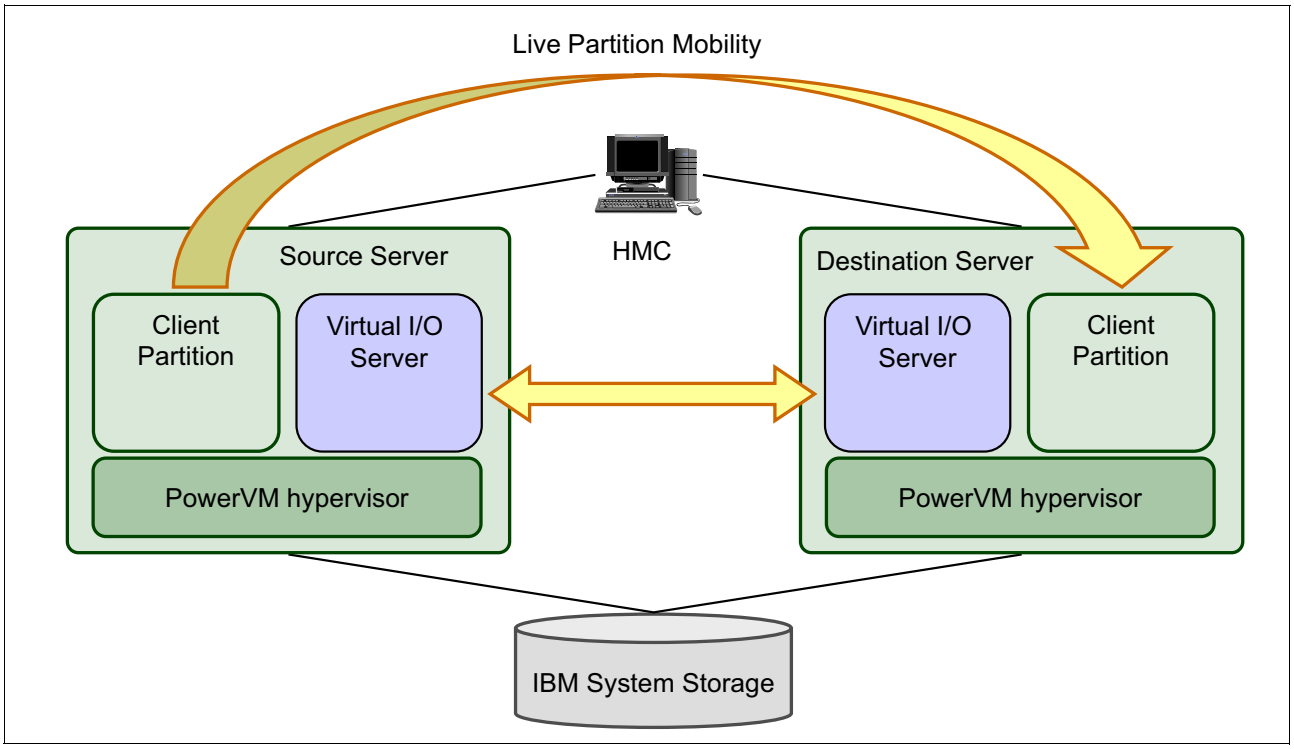


Figure 2-14 An example of Live Partition Mobility

An environment that has only small windows for scheduled downtime might use LPM to manage many scheduled activities either to reduce downtime through inactive migration or to avoid service interruption through active migration.

For example, if a system must be shut down due to a scheduled maintenance, its partitions can be migrated to other systems before the outage.

An example is shown in Figure 2-15, where system A must be shut down. The production database partition is actively migrated to system B, and the production web application partition is actively migrated to system C. The test environment is not considered vital, so it is shut down during the outage.

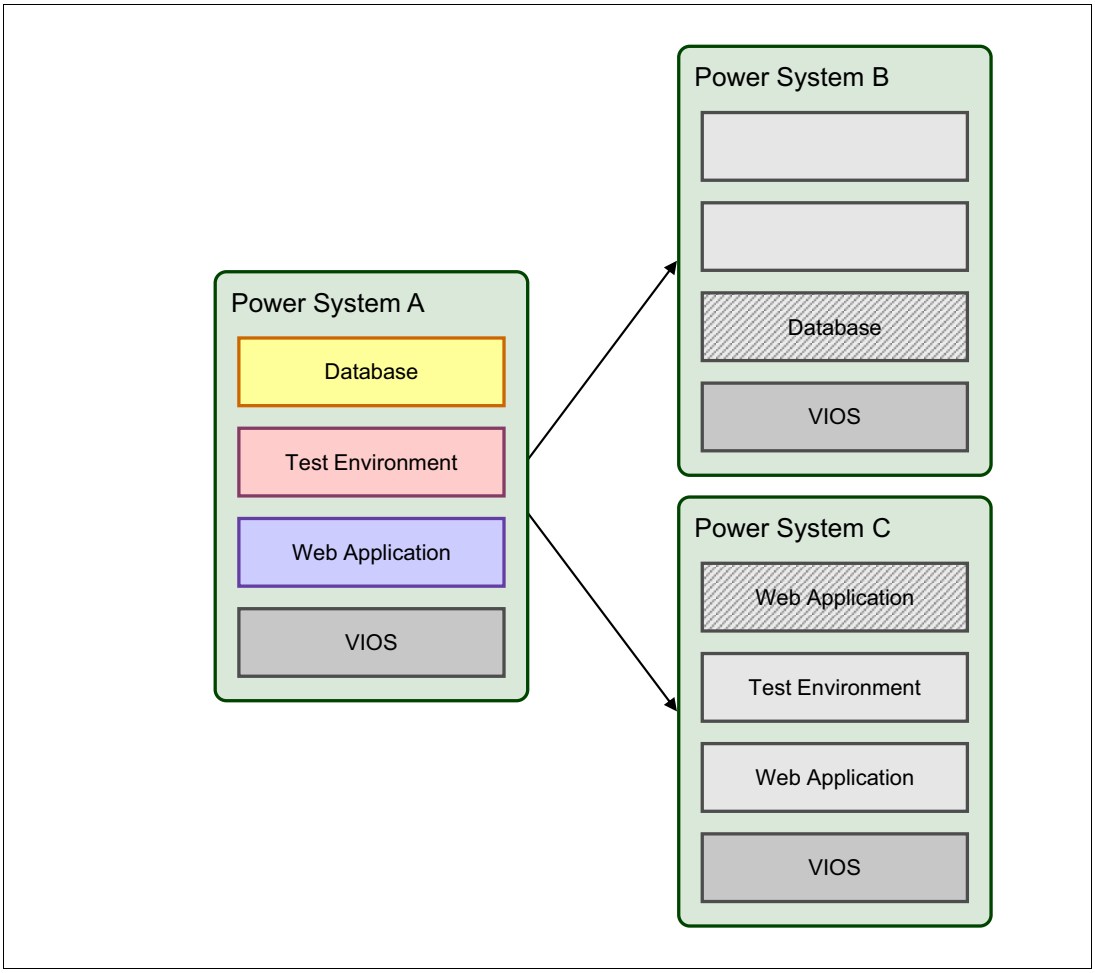


Figure 2-15 Migrating all partitions of a system

LPM is a reliable procedure for system reconfiguration and might be used to improve the overall system availability. LPM increases global availability, but it is not a high availability (HA) solution to avoid an unplanned outage. It requires both source and destination systems to be operational and that the partition is not in a failed state. In addition, it does not monitor operating system and application states, and it is by default a user-initiated action.

Scenario: LPM can help to reduce only a planned outage.

Benefits of partition mobility

Partition mobility provides systems management flexibility and improves system availability as follows:

- Avoid planned outages for hardware or firmware maintenance by moving LPARs to another server and then doing the maintenance. LPM can help lead to zero downtime maintenance because you can use it to work around scheduled maintenance activities.
- Avoid downtime for a server upgrade by moving LPARs to another server and then doing the upgrade. This approach allows your users to continue their work without disruption.

- Conduct preventive failure management. If a server indicates a potential failure, you can move its LPARs to another server before the failure occurs. Partition mobility can help avoid unplanned downtime.
- Optimize server workloads:
  - Workload consolidation

    You can consolidate workloads that run on several small, underutilized servers onto a single large server.
  - Flexible workload management

    You can move workloads from server to server to optimize resource use and workload performance within your computing environment. With active partition mobility, you can manage workloads with minimal downtime.

The migration manager function is on the HMC and in charge of reconfiguring source and target systems. It checks that all hardware and software prerequisites are met. It runs the required commands on the two systems to complete migration while providing the migration status to the user.

When an inactive migration is performed, the HMC invokes the configuration changes on the two systems. During an active migration, the running state (memory, registers, and so on) of the mobile partition is transferred during the process.

Memory management of an active migration is assigned to a mover service partition (MSP) on each system. During an active partition migration, the source MSP extracts the mobile partition's state from the source system. Then, it sends partition's state over the network to the destination MSP, which updates the memory state on the destination system.

LPM has no specific requirements on the mobile partition's memory size or the type of network that connects the MSPs. The memory transfer is a process that does not interrupt a mobile partition's activity and might take time when a large memory configuration is involved on a slow network. Use a high-bandwidth connection, such as 1 Gbps Ethernet or larger.

While partition mobility provides many benefits, it does not perform the following functions:

- Partition mobility does not provide automatic workload balancing.
- Partition mobility does not provide a bridge to new functions. LPARs must be restarted and possibly reinstalled to leverage new features.

When an LPAR is moved by using LPM, a profile is automatically created on the target server that matches the profile on the source server. Then, the partition’s memory is copied asynchronously from the source system to the target server, which creates a clone of a running partition. Memory pages that changed on the partition (“dirty” pages) are recopied. When a threshold is reached that indicates that enough memory pages were successfully copied to the target server, the LPAR on that target server becomes active, and any remaining memory pages are copied synchronously. Then, the original source LPAR automatically is removed.

Because the HMC always migrates the last activated profile, an inactive LPAR that was never activated cannot be migrated. For inactive partition mobility, you can either select the partition state that is defined in the hypervisor or select the configuration data that is defined in the last activated profile on the source server.

For more information about LPM, see the following resources:

- Partition mobility, found at:

  https://www.ibm.com/docs/en/power10/9080-HEX?topic=environment-live-partition-mobility

- *Live Partition Mobility Preparation Checklist*, TIPS1185
- Preparing for partition mobility, found at:

  https://www.ibm.com/docs/en/power10/9105-41B?topic=mobility-preparing-partition

- *Live Partition Mobility Setup Checklist*, TIPS1184
- *Implementing IBM VM Recovery Manager for IBM Power Systems*, SG24-8426

What is new in Live Partition Mobility

Some of the most recent updates to LPM include the following items:

- Virtual software tier capability support and vNIC auto-priority fail-over support on the destination server.
- Changes in the processor compatibility mode.
- Enhancements to the HMC GUI.
- Both Active and Inactive Partition Mobility follow a support statement of N-2 releases.

  Power10 processor-based servers support migration between Power10, Power9, and Power8 processor-based servers, but Power10 processor-based servers do not support direct migration from POWER7 processor-based servers.

- LPM and SRR succeed even if optical devices are configured with no media that are installed in the device.

  Previously, if optical devices were configured, LPM and Remote Restart failed.

- HMC determines the fastest adapter that is available in the MSP automatically unless the user specifies to migrate by using a specific adapter.

  Previously, LPM did not consider adapter speed, type of connection (direct or SEA), and existing LPM traffic on possible MSP connections.

For more information about these updates and new or changed information and capabilities in LPM since the previous update, see the following resources:

- What's new in Live Partition Mobility, found at:

  https://www.ibm.com/docs/en/power10/9080-HEX?topic=mobility-whats-new-in-live-partition

- *Power10 PowerVM Overview*, found at:

  https://public.dhe.ibm.com/systems/power/community/aix/PowerVM_webinars/110_P10_PowerVM.pdf

- PowerVM features in the new Power10 Servers, found at:

  https://community.ibm.com/community/user/power/blogs/pete-heyrman1/2021/09/27/powervm-features-in-the-new-power10-servers

NPIV LUN or disk level validation on VIOS for a partition mobility environment

By default, partition mobility validation for NPIV devices is checked only up to the port level. This approach might result in client failures if the actual disks that are mapped to the client on the source system are not masked to be accessed by using inactive WWPNs. Along with the port-level validation, it is also possible to validate up to the disk mapping.

Consider the scenario where the partition mobility operation for an NPIV client completed successfully but the client lost disk I/O to the SAN storage immediately after the migration completed. This situation commonly occurs when the SAN zoning is correct but the storage is not provisioned to the mobile partition's inactive WWPNs, which are used by the mobility operation. Enabling NPIV LUN or disk-level validation on VIOS might help in this situation.

Disk validation can add a considerable amount of time to partition mobility validation for clients that are using NPIV disks. The amount of time that is required to validate NPIV devices up to the disk level depends on the number of disks that are mapped to a client. For larger configurations, the extra time that is spent in validation might have a noticeable impact on the overall time that is required to migrate the partition. Therefore, as a best practice, consider performing periodic partition mobility validation with LUN level validation enabled. Also, consider planning the validation outside of scheduled maintenance windows. Either skip validation, or run validation with LUN level validation disabled when partition mobility operations must be completed in a short period.

For more information, see NPIV LUN or disk level validation, found at:

https://www.ibm.com/docs/en/power10/9080-HEX?topic=pseudodevice-npiv-lun-disk-level-validation

2.7 Simplified Remote Restart

When a server crashes, the partitions on that server also crash. To mitigate the impact of a server outage, you can use *SRR*, which is a PowerVM HA option for LPARs. When an error causes a server outage, a partition that is configured with the SRR capability can be restarted on a different physical server. Sometimes, recovering from a server error takes a long time, in which case the SRR feature can be used for faster reprovisioning of the partition. This operation completes faster compared to recovering the failed server and then restarting the partition.

Here are the characteristics of the SRR feature:

- During the SRR operation, the LPAR is shut down and then restarted on a different system.
- The SRR feature preserves the resource configuration of the partition. If processors, memory, or I/O are added or removed while the partition is running, the SRR operation activates the partition with the most recent configuration.

For more information, see Simplified Remote Restart, found at:

https://www.ibm.com/docs/en/power10/9080-HEX?topic=9080-HEX/p10eew/p10eew_remmres.htm

Here is a typical use case for SRR:

- The user creates a partition with SRR capability on a capable server. We call this server the *source server*.
- The user can enable and disable the capability anytime after the partition is created and toggle support for the capability.
- The user assigns resources to the partition. The resource restrictions are similar to LPM, that is, no dedicated I/O, no Host Ethernet Adapter (HEA) adapter, no Host Channel Adapter (HCA) adapter, no OptiConnect, no server adapter in IBM i partition, and so on.
- Storage attached to the partition through virtual I/O should be accessible from another server (similar to LPM). We call this server the *destination server.*
- When the user activates the partition, the configuration of the partition along with partition state data are collected and persisted automatically on the HMC.
- The data that is persisted on the HMC is updated automatically for any configuration change.
- When the source server crashes, the user initiates the remote restart operation to restart the partition on the destination server.
- After the source server is back to operating state, the HMC runs an automatic cleanup of the partition, which is now restarted on a new (destination) server.

Figure 2-16 shows the SRR configuration setup.

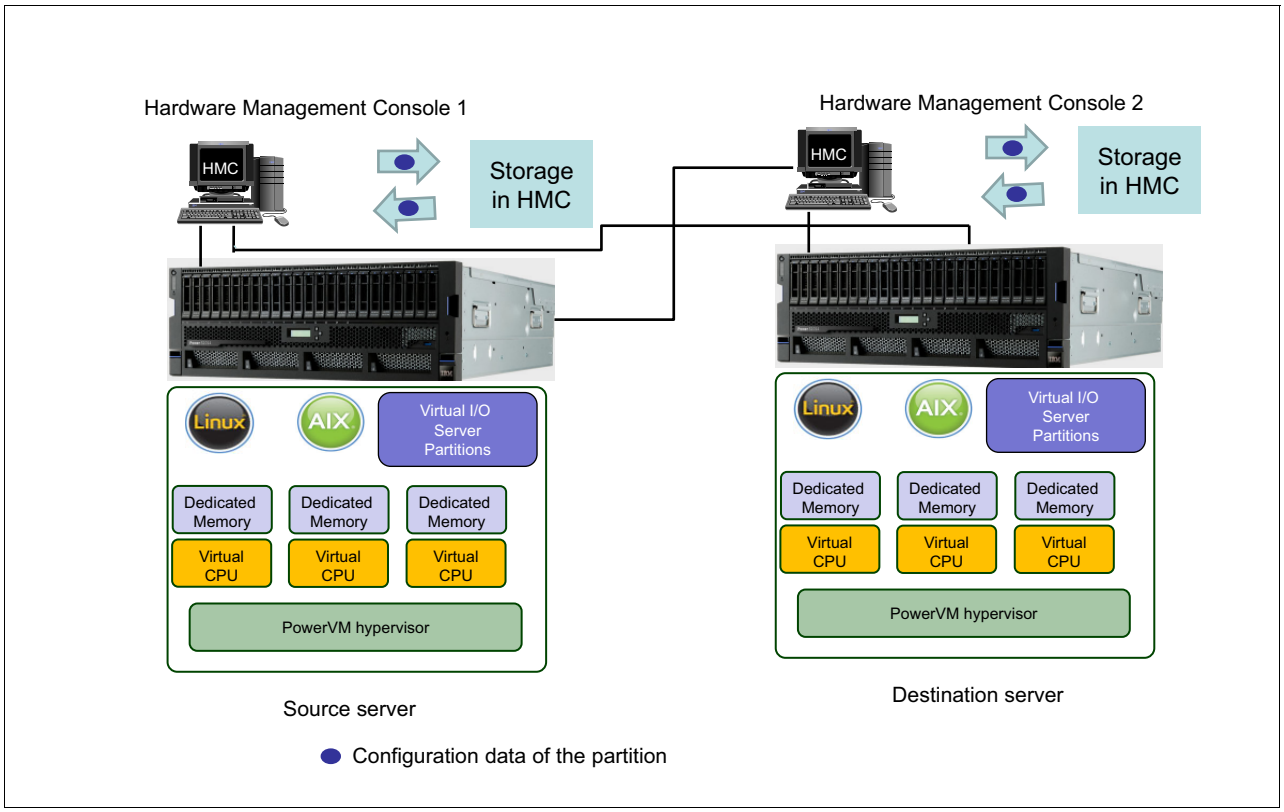


Figure 2-16 Simplified Remote Restart configuration setup

For an SRR-capable partition, configuration data is stored on the HMC. Each HMC managing the server persists its own copy of the configuration information. Any storage that is used by the partition must be through vSCSI or VFC Storage that is accessible from both the source and destination servers. In general, the configuration requirements and restrictions are like the ones for LPM.

When an LPAR is restarted by using SRR, a new partition is created automatically on the target server that matches the partition on the source server. Then, the new partition is mapped to the storage LUNs that were being used by the original partition, which is now inactive again. Then, the new partition on the target server is activated and the partition becomes active. When the source server becomes active, the HMC automatically initiates a cleanup of the partition on the source server.

For more information, see the following resources:

- Simplified Remote Restart, found at:

  https://www.ibm.com/docs/en/power10/9786-42H?topic=9786-42H/p10eew/p10eew_remmres.html

- *Implementing IBM VM Recovery Manager for IBM Power Systems*, SG24-8426

Enhanced Simplified Remote Restart capabilities

Many SRR enhancements were introduced after its initial release:

- Remote restart a partition with reduced or minimum CPU or memory on the target system.

  When a system outage occurs and you want to restart all the partitions on another system to reduce the downtime, in some scenarios the target system does not have enough capacity to host all the partitions. However, some partitions might be run with reduced resources (for example, development or test workloads versus production workloads). You now can restart a partition with reduced resources on the target system.

- Remote restart by choosing a different virtual switch on the target system.

  You can validate or perform the remote restart operation for an LPAR when you want to start the LPAR with a different virtual switch on the target server than the virtual switch that the LPAR was assigned on the source server.

- Remote restart the partition without powering on the partition on the target system.

  With this option, you can prevent an LPAR from being started during the remote restart operation. You can use this option in cases where you want to check the configuration on the target system before the partition is powered on. All the other steps that are performed during a remote restart operation are performed in this case except for powering on the partition.

- Remote restart the partition for test purposes when the source managed system is in the Operating or Standby state.

  Remote restart is supported when a system fails. If you want to validate whether a remote restart operation works in a system failure, you always can use the validate option. However, if you want to go one step further and test partition restart on another system, you can do that by using the test option in the HMC. The source partition *must be in the shutdown state* to use the test option for remote restart.

- Display the partition configuration information.

  HMC collects and persists the configuration data that is required for restarting a partition. You can now use the **`lsrrstartlpar`** command to view the persisted configuration information of all the LPARs that support SRR.

- Remote restart by using the REST API.

  You can also run the remote restart operation by using the Representational State Transfer (REST) API:

  `https://<<HMCIP>>:12443/rest/api/uom/ManagedSystem/<ManagedSystem_UUID>LogicalPartition/<<PARTITION_UUID>>/do/RemoteRestart`

  All the options and overrides are added to the REST API too.
- IBM Power Virtualization Center (PowerVC) can initiate a single partition restart or a whole frame restart, but not a subset of SRR-capable partitions.

2.7.1 PowerVC automated remote restart

You can use PowerVC to remotely restart virtual machines (VMs) from a failed host (source host) to another host (destination host). After the failed host is restarted, any VMs that were remotely restarted are automatically removed from that host, but they remain on the destination host. The count of VMs is updated on the source host when it is restarted.

Note: If the VM was shut down on the source host, it remains shut down on the target host after the remote restart.

You can define on the PowerVC user interface whether the VMs on a host must be remote-restarted by using a manual process or an automated process.

Automated remote restart (ARR) monitors hosts for failure by using the Platform Resource Scheduler (PRS) HA service. If a host fails, PowerVC automatically remote restarts the VMs from the failed host to another host within a host group.

Without ARR enabled, when a host goes into the `Error` or `Down` state, you must manually trigger the remote restart operation. You can manually remote restart VMs from a host at any time, regardless of its ARR setting.

For more information, see *r*emote restart virtual machines from a failed host, found at:

`https://www.ibm.com/docs/en/powervc/2.0.3?topic=restart-remote-virtual-machines-from-failed-host#powervc_vm_recovery_hmc`

Note: By default, the ARR feature is disabled so that the administrator can select the host that will be considered for ARR.

For more information, see Remote restart, found at:

`https://www.ibm.com/docs/en/powervc/2.0.3?topic=remote-restart`

2.8 VM Recovery Manager

Organizations that use Power servers might need simple and economical high availability and disaster recovery (HADR) solutions based on PowerVM for AIX, IBM i, and Linux environments.

IBM VM Recovery Manager (VMRM) for Power Systems is an HADR solution that enables VMs to be moved between systems by using LPM for planned events or restarted on another system for unplanned outages. VMs also can be replicated and restarted at remote locations for DR operations, and DR testing can be done with a single command.

Unlike traditional clustering solutions that are operating system-aware, VMRM is a solution that runs regardless of the operating systems, which enables a uniform HADR solution across AIX, IBM i, and Linux environments.

VMRM is an active/inactive configuration where the production VMs can be moved to a cold standby configuration by either LPM or a VM restart procedure. The active/inactive configuration means that software licenses are not required on the target system because the production applications and operating system are restarted onto the target system.

VMRM comes in two versions:

- VM Recovery Manager HA (VMRM HA)
- VM Recovery Manager DR (VMRM DR), formerly called Geographically Dispersed Resiliency (GDR)

VMRM HA (5765-VRM) provides the VM restart-based HA management, and VMRM DR (5765-DRG) provides VM restart-based HA or DR capabilities.

2.8.1 VM Recovery Manager HA

HA management is a critical feature of business continuity plans. Any downtime to the software stack can result in loss of revenues and disruption of services. IBM VMRM HA for Power Systems is an HA solution that is easy to deploy and provides an automated solution to recover the VMs (LPARs).

The VMRM HA solution implements recovery of the VMs based on the VM restart technology. The VM restart technology relies on an out-of-band monitoring and management component that restarts the VMs on another server when the host infrastructure fails. The VM restart technology is different from the conventional cluster-based technology that deploys redundant hardware and software components for a near-real-time fail-over operation when a component fails. The cluster-based HA solutions are commonly deployed to protect critical workloads.

The VMRM HA solution is ideal to ensure HA for many VMs. Additionally, the VMRM HA solution is easier to manage because it does not have clustering complexities.

Figure 2-17 shows the failed VMs that are restarted on an adjacent server. It is a shared storage topology, which is used in along with LPM.

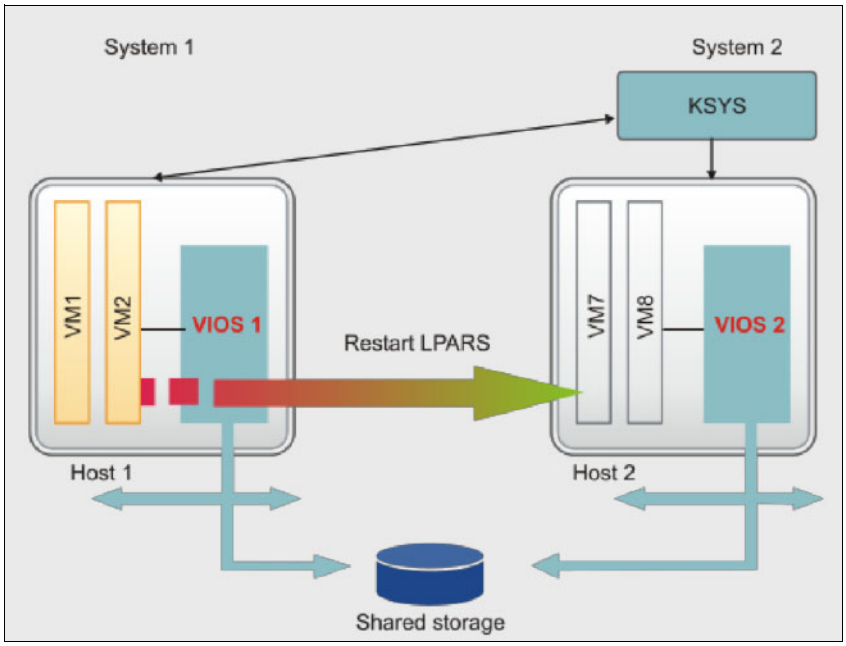


Figure 2-17 VMRM HA: Failed VMs restarted on adjacent server

- The secondary server or partition is inactive until VMs are restarted on it or if LPM is used for a planned outage event.
- There is one physical copy and one logical copy (unless deployed in an IBM HyperSwap® type of configuration in which case there are two physical copies of the data).

The VMRM HA solution provides the following capabilities:

- Host health monitoring
- Unplanned HA management
- Planned HA management
- Advanced HA policies
- GUI- and CLI-based management

For more information about these capabilities, see VM Recovery Manager HA overview, found at:

https://www.ibm.com/docs/en/vmrmha/1.6?topic=overview

2.8.2 VM Recovery Manager DR

DR of applications and services is a key component to provide continuous business services. The IBM VMRM DR for Power Systems solution is a DR solution that is easy to deploy and provides automated operations to recover the production site. The VMRM DR solution is based on the IBM Geographically Dispersed Parallel Sysplex® (IBM GDPS®) offering concept that optimizes the usage of resources. This solution does not require you to deploy the backup VMs for DR. Thus, the VMRM DR solution reduces the software license and administrative costs.

Note: VMRM DR now supports the SR-IOV vNIC.

The following HA and DR models are commonly used by organizations:

- Cluster-based technology
- VM restart-based technology

Figure 2-18 shows the DR solution that uses the VM restart-based technology. The VMRM DR solution uses this model.

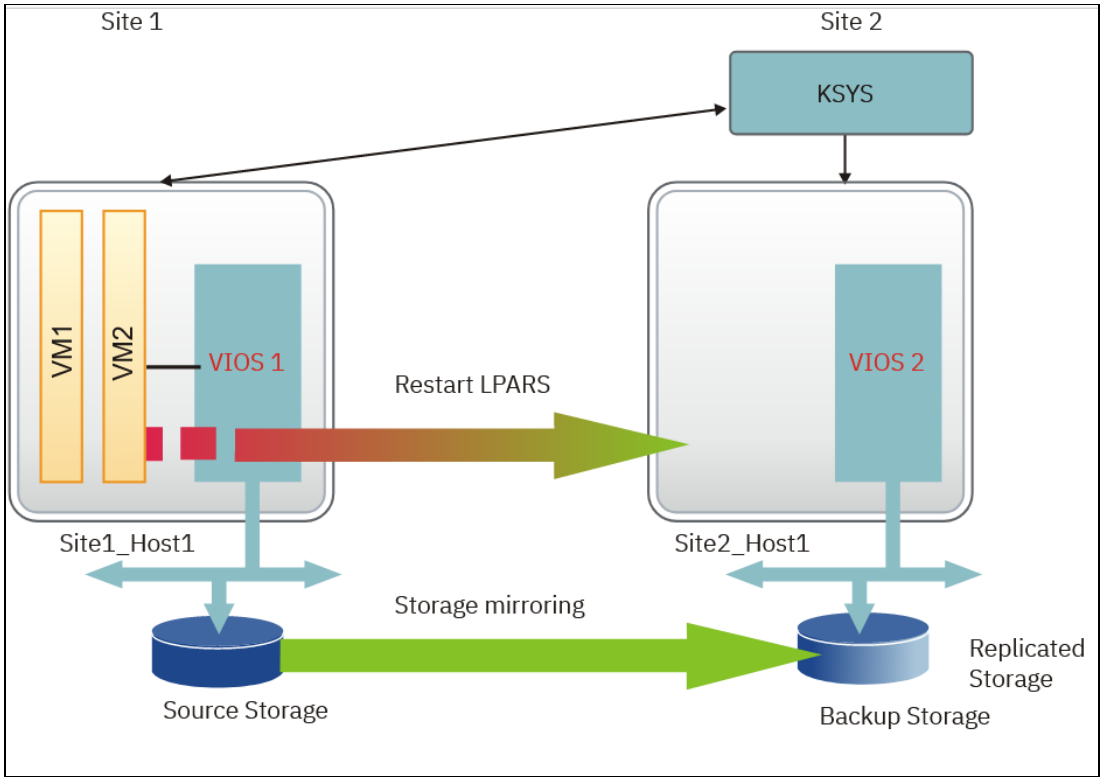


Figure 2-18 VMRM DR: VM restart-based disaster recovery model

VMRM DR is based on the replication of the VMs to a remote location:

- The DR server is inactive until the replicated VMs are restarted on it.
- If the production system fails (or will be tested for DR compliance), the VMs are restarted on a secondary system in the cluster.
- There are two physical copies of the VMs and one logical copy in this configuration.

VMRM for DR includes the capability to manage both restart HADR from a single KSYS simultaneously, as shown in Figure 2-19.

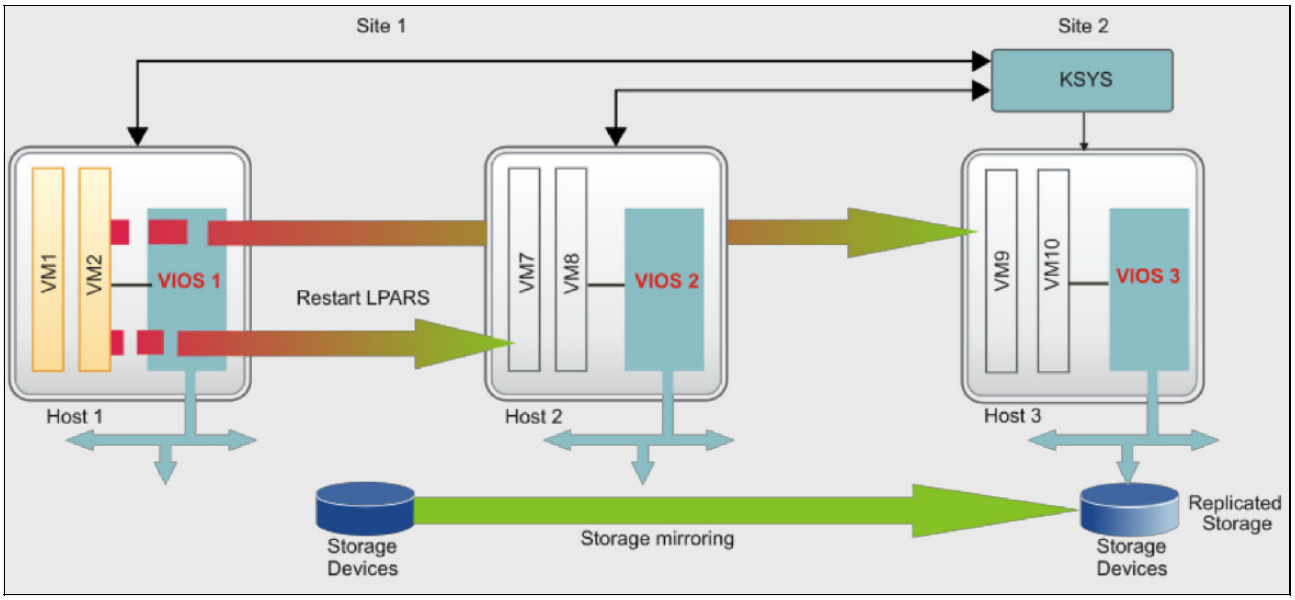


Figure 2-19 VM Recovery Manager DR combines HA and DR

VMRM for DR 1.5 enables VMRM HA at both the primary and secondary sites.

For more information, see Overview for IBM VM Recovery Manager DR for Power Systems, found at:

https://www.ibm.com/docs/en/vmrmdr/1.6?topic=overview

2.9 Capacity on Demand

CoD offerings allow users to dynamically activate one or more resources on your server as required by the workload. CoD allows users to activate inactive processors or memory units that are installed on your server on a temporary or permanent basis.

The following terminology is used in CoD:

Active Resources Active processor cores and memory that are available for use on the server.

Inactive Resources Inactive processor cores and inactive memory units are resources that are included with your server but are not available for use until you activate them.

2.9.1 CoD offerings

Depending on your workload requirements, you can choose from the available CoD offerings. A brief description of each offering is in the following sections.

Trial Capacity

Some of the main characteristics of Trial Capacity are as follows:

- Users can evaluate the use of inactive processor cores, memory, or both at no charge with Trial Capacity.
- Trial Capacity provides a no-charge temporary activation of resources. This feature helps users to test new capabilities and functions that are available on Power servers.
- Trial Capacity provides 30 days activation of resources. The trial period advances only when the server is powered on.
- When an action is required after the implementation, HMC displays messages to notify the user.
- Users can stop a current Trial Capacity for processor cores or memory units before the trial automatically expires by using the HMC.
- If a user chooses to stop the trial before it expires, the user cannot restart the Trial Capacity for any remaining days.
- There are two types of trials: standard and exception.

For more information, see IBM Entitled Systems Support (IBM ESS), found at:

https://www.ibm.com/servers/eserver/ess

Capacity Upgrade on Demand (Permanent Activation)

Some of the main characteristics of Capacity Upgrade on Demand (CUoD) are as follows:

- CUoD allows users to activate permanently inactive processor cores and memory units. This offering helps users to meet new workload or future growth requirements.
- Users can order activation features for a new server or an installed server. After the order is placed, the user receives a code that activates inactive processor cores or memory units.
- For a new server, the order can contain one or more activation features for processor cores or memory units, which result in one or more activation codes. In this case, the activation codes are entered before the server is shipped.
- When users order CUoD activation features for an installed server, the users must determine whether they want to permanently activate some or all their inactive processor cores or memory units.
- Users must order one or more activation features and then use the resulting activation codes to activate their inactive processor cores or memory units.

Elastic Capacity (4586-COD)

Some of the main characteristics of Elastic Capacity are as follows:

- With Elastic Capacity, users can activate processor cores or memory units for several days by using the HMC to activate resources on a temporary basis.
- With Elastic Capacity features, users can purchase Elastic Capacity for processor and memory days that temporarily activate processors and memory in full-day increments.

- With this new implementation, users can add temporary processor and memory capacity whenever needed in minutes without any intervention from IBM.
- The procedure to activate Elastic Capacity can be summarized as follows:
  - Elastic Capacity for processor and memory days can be purchased from IBM sellers or IBM Business Partners.
  - Elastic Capacity days can be applied to any similar CoD-enabled IBM Power server that is in your enterprise.
  - After the days are ordered through the IBM seller or IBM Business Partner, they become available in IBM ESS for users to provision to selected servers and download activation codes in minutes.

These features provide the following benefits:

- A quick and convenient way to order Elastic Capacity for processor and memory days through your IBM seller or IBM Business Partner and get billed at the end of the month.
- The ability to order in bulk under one order and allocate the processors and memory to multiple systems in the same country.
- No need to report usage monthly.
- The ability to order and provision in minutes.

With these Elastic Capacity features, system administrators can download Elastic Capacity activation codes for a specific number of processor or gigabyte memory days. Then, they enter the activation codes directly at the HMC to activate resources.

The HMC logs the usage activity and notifies you when you are running low on Elastic Capacity for processor and memory days, at which time you can provision more or purchase extra days. Before you use temporary capacity on your server, ensure that the minimum required firmware SP is installed.

The Elastic Capacity offering is available to use for the following IBM Power servers only:

- Supported IBM Power midrange systems: 9043-MRX, 9040-MR9, 8408-44E, and 8408-E8E
- Supported IBM Power high-end systems: 9080-HEX, 9080-M9S, 9080-MHE, 9080-MME, 9119-MHE, and 9119-MME

For more information, see Elastic Capacity on Demand, found at:

https://www.ibm.com/docs/en/power10?topic=demand-elastic-capacity

IBM Power Enterprise Pools

Power Enterprise Pools (PEP) is a solution that is built on top of CoD features. PEP allows a group of systems to share their processor and memory activations. For more information, see 2.10, “Power Enterprise Pools” on page 70.

CoD offerings summary

Table 2-1, Table 2-2, and Table 2-3 provide a summary of available CoD offerings.

Table 2-1 Enterprise Power servers support

CoD offerings	Duration	Power E980	Power E1080
Trial Capacity	30 days	Yes	Yes
Permanent Activation	Permanent	Yes	Yes
Elastic Capacity - 4586-COD	1 day or more	Yes	Yes
Enterprise Pools 2.0 - 5819-CRD	1 minute or more	Yes	Yes

Table 2-2 Midrange Power servers support

CoD offerings	Duration	Power E950	Power E1050
Trial Capacity	30 days	Yes	Yes
Permanent Activation	Permanent	Yes	Yes
Elastic Capacity - 4586-COD	1 day or more	Yes	Yes
Enterprise Pools 2.0 - 5819-CRD	1 minute or more	Yes	Yes

Table 2-3 Scale-out Power servers support

CoD offerings	Duration	Power S922 Power S924	Power S1022 and Power S1024
Trial Capacity	30 days	N/A	N/A
Permanent Activation	Permanent	N/A	YES
Elastic Capacity - 4586-COD	1 day or more	N/A	N/A
Enterprise Pools 2.0 - 5819-CRD	1 minute or more	Yes[a]	Yes[a]

a. 5819-CRD is not applicable for memory on scale-out Power servers because all their installed memory must be fully activated by using static licenses.

CoD Management by using HMC

HMC provides an easy-to-use interface to accomplish the following tasks:

- View available CoD resources.
- Enable and disable CoD resources.
- View the CoD history log.

Figure 2-20 provides an overview of the CoD functions in the HMC.

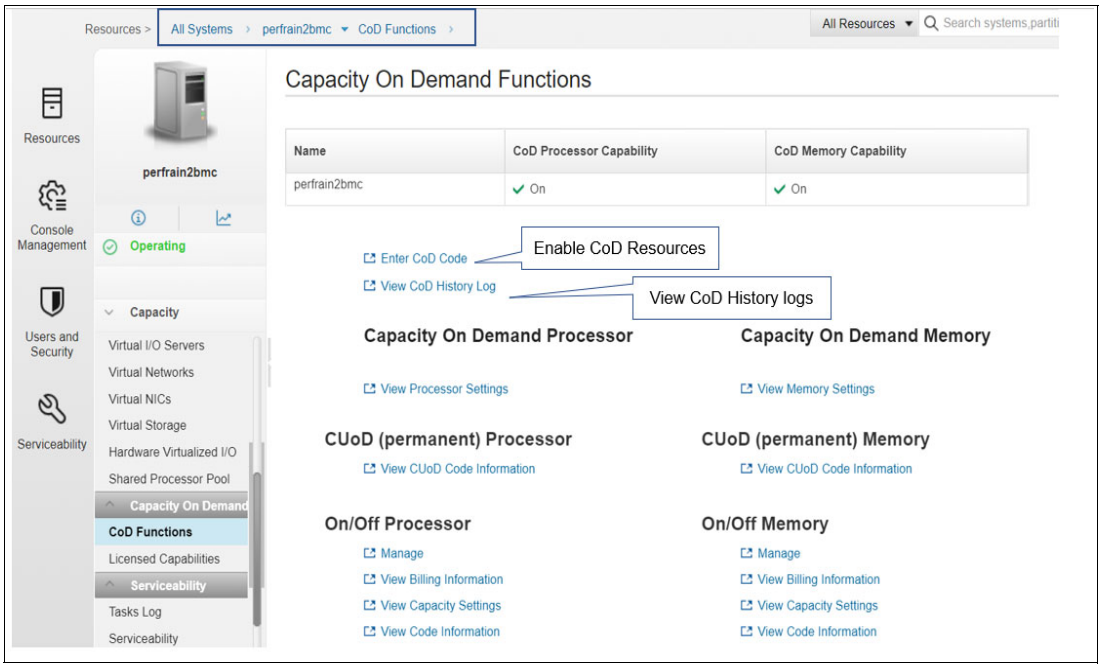


Figure 2-20 CoD management by using HMC

CoD advanced functions

CoD advanced functions can be used to activate the following capabilities:

- Enterprise Enablement features

  Enterprise Enablement is a CoD advanced function technology that enables the system for 5250 online transaction processing (OLTP).

- AME

  For more information about AME, see 2.2.2, “Active Memory Expansion” on page 39.

- WWPN renewal code

  Power servers that use NPIV support up to 64,000 unique WWPNs. Servers with 64,000 customers must request a WWPN renewal code on the website. The code enables a WWPN prefix, which provides the first 48 bits of each WWPN and makes 64,000 more WWPNs available on the server.

For more information, see Capacity on Demand, found at:

https://www.ibm.com/docs/en/power10/9080-HEX?topic=environment-capacity-demand

2.10 Power Enterprise Pools

The PEP feature enables a group of systems to work together as a pool of resources, which provides resource flexibility and cost efficiency.

Two different PEP features are available:

- Power Enterprise Pools 1.0
- Power Enterprise Pools 2.0 (IBM Power Systems Private Cloud with Shared Utility Capacity)

2.10.1 Power Enterprise Pools 1.0

The PEP 1.0 feature is the first version of PEP. With PEP 1.0, servers can be configured with Static and Mobile Activations. After the servers are added to a pool, a system administrator can move Mobile Activations in between machines by using the HMC to activate resources where they are needed. This capability allows customers to balance processor and memory resources across the pool of machines and also provides cost efficiency by reducing the total number of activation requirements across systems and data centers. PEP 1.0 is available only on specific enterprise class Power servers. It allows pooling of two consecutive generations of hardware together in the same pool. At the time of writing, there are four possible combinations:

- An IBM Power 770+, IBM Power E870, IBM Power E870C, and IBM Power E880C pool
- An IBM Power 780+, IBM Power 795, IBM Power E880, E870C, and IBM Power E880C pool
- An IBM Power E870, IBM Power E880, IBM Power E870C, IBM Power E880C, and IBM Power E980 pool
- An IBM Power E980 and IBM Power E1080 pool

For more information about PEP 1.0, see Power Enterprise Pool, found at:

https://www.ibm.com/docs/en/power10?topic=demand-power-enterprise-pool

2.10.2 Power Enterprise Pools 2.0 (IBM Power Systems Private Cloud with Shared Utility Capacity)

PEP 2.0, also known as IBM Power Systems Private Cloud with Shared Utility Capacity, is another licensing model and it is different from PEP 1.0. In PEP 2.0, Power servers are configured with Base Activations. When the machines are first installed, Base Activations determine the amount of physical capacity of the server to be activated, and the rest of the physical capacity stays inactive. When a machine is added to a PEP 2.0 pool, all installed processors and memory resources are automatically activated and made available for immediate use. The machine's Base Activations are added to the pool's base capacity.

All usage under the base capacity across the pool is not charged. All usage that is above the base capacity is metered by the minute and charged on a pay-as-you-go basis. Memory usage is tracked by assignment to active VMs, and core usage is tracked by the actual usage of VMs in the pool.

A single pool can contain up to 64 Power servers. These servers must be in the same enterprise and country. Up to 1,000 VMs can be supported by a single HMC. A single IBM Cloud Management Console (IBM CMC) instance can support up to 4,000 VMs either in one large pool or in several smaller pools.

Power E1080, Power E1050, Power E980, Power E950, Power S924 (9009-42G), and Power S922 (9009-22G), Power S1022 (9105-22A), and Power S1024 (9105-42A) servers are supported in PEP 2.0.

The following types of pools are available in PEP 2.0:

- A Power E980 and Power E1080 pool
- A Power E950 and Power E1050 pool
- A Power S922, Power S924, Power S1022, and Power S1024 pool

PEP 2.0 is configured, managed, and monitored by using IBM CMC. IBM CMC plays an important role in an IBM Power Systems Private Cloud with Shared Utility Capacity solution. Figure 2-21 shows a solution diagram for IBM CMC and PEP 2.0.

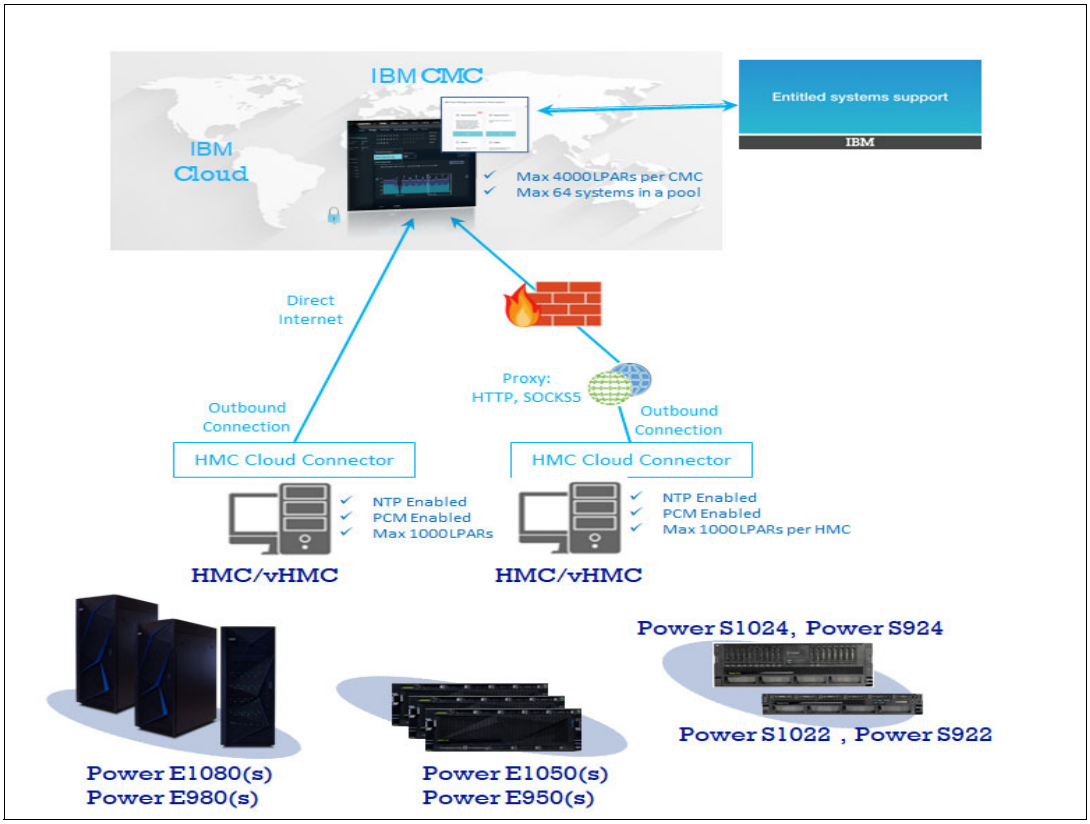


Figure 2-21 IBM Cloud Management Console solution diagram

2.10.3 Comparing PEP 1.0 and PEP 2.0

PEP 1.0 and PEP 2.0 are both built by using CoD features, but they are fundamentally different solutions. PEP 1.0 is built on top of Static and Mobile Activations where Mobile Activations can be moved in between pool member machines by using HMC. In PEP 2.0, all the hardware resources are by default activated and made available for consumption, and then the utilization is tracked by the minute by IBM CMC. In other words, in PEP 1.0, resources must be activated and made ready before you can use them, but in PEP 2.0 all the resources are activated, and the customer is charged for only their actual utilization.

Elastic Capacity can be used to activate extra resources in PEP 1.0, but in PEP 2.0 all the resources already are activated. In PEP 2.0, there are no extra resources to activate by using CoD features.

PEP 1.0 is available only for enterprise class Power server models. PEP 2.0 is available for scale-out, mid-range, and scale-up server models.

Table 2-4 on page 73 provides a further comparison between PEP 1.0 and PEP 2.0.

Table 2-4 Comparing PEP 1.0 and PEP 2.0

Feature / Attribute	PEP 1.0 with Mobile Capacity	PEP 2.0 with Utility Capacity
Systems supported.	Power 770, Power 780, Power E870/E870C, Power E880/E880C, Power E980, and Power E1080	Power E1080, Power E980, Power E1050, Power E950, Power S922 and Power S924 (G Models), Power S1022 and Power S1024
Interoperability.	► Power 770+, Power E870, Power E870C and Power E880C ► Power 780+, Power 795, Power E880, Power E870C, and Power E880C ► Power E870, Power E880, Power E870C, Power E880C, and ► Power E980 ► Power E980 and Power E1080	► Power E1080 and Power E980 ► Power E1050 and Power E950 ► Power S922, Power S924 (G Models), Power S1022, and Power S1024
Activations supported.	Static, Mobile-enabled, and Mobile.	Base.
Purchased Capacity sharing.	Mobile: Shared manually or by using scripts.	Base: Pooled and effectively shared seamlessly without intervention.
Variable usage.	Elastic Capacity: Allocated per system, enabled for use manually or by using scripts, and charged by the day. Utility Capacity (processor) by the minute (except Power10 processor-based servers).	Metered: Always active, used automatically when needed, and charged by the minute if the pool usage is above its Base Capacity.
Resource usage tracking.	May be collected by the system and aggregated manually or programmatically.	Collected and aggregated across a pool and available by resource, VM, and system historically by using IBM CMC.
Management Interface	HMC.	IBM CMC.
CoD key, XML, and codes management.	HMC Config file (XML) manually applied by using HMC.	All enablement keys are applied transparently by using IBM CMC or the HMC. No intervention is required.
Add or remove systems from a pool.	Manually by using a supplement.	Self-service by using IBM CMC.
Budgeting support.	Elastic Capacity processor and memory days may be allocated to a system by an administrator and tracked manually.	Yes, monthly by pool, with automated enforcement and tailorable alerts.
Processor partitions supported.	Dedicated processor partitions. Shared processor partitions.	Dedicated Processor Partitions Shared Processor Partitions
Supports multiple HMCs per system.	Yes.	Yes.

Feature / Attribute	PEP 1.0 with Mobile Capacity	PEP 2.0 with Utility Capacity
Requires connection to IBM Cloud.	No.	Yes.
Number of systems that are supported in a single pool.	N/A.	64.
Cross-border pools allowed.	Single country and within the European Union Community.	Single country.
Number of LPARs.	No limit.	4000.
Minimum resource activation required.	Minimum of eight cores of Static Processor Activation except for the Power E1080, which requires a minimum of 16 Static Activations. 50% memory (100 GB increments).	One core per machine (all Base shared). 256 GB of memory (256 GB increments).

2.10.4 Migrating from PEP 1.0 to PEP 2.0

It is possible to migrate a server from a PEP 1.0 configuration to PEP 2.0. However, by design, a Power server cannot be a part of two different pools concurrently. Therefore, systems must be removed from the PEP 1.0 pool before they are added to the PEP 2.0 pool. This step is followed by a miscellaneous equipment specification upgrade where PEP 1.0 markers are replaced with PEP 2.0 markers, and Static and Mobile Activations are converted to Base Activations. After the conversions are done and reflected on the IBM ESS website, the machine can participate in a PEP 2.0 pool.

Figure 2-22 on page 75 shows a flow diagram for conversion from PEP 1.0 to PEP 2.0.

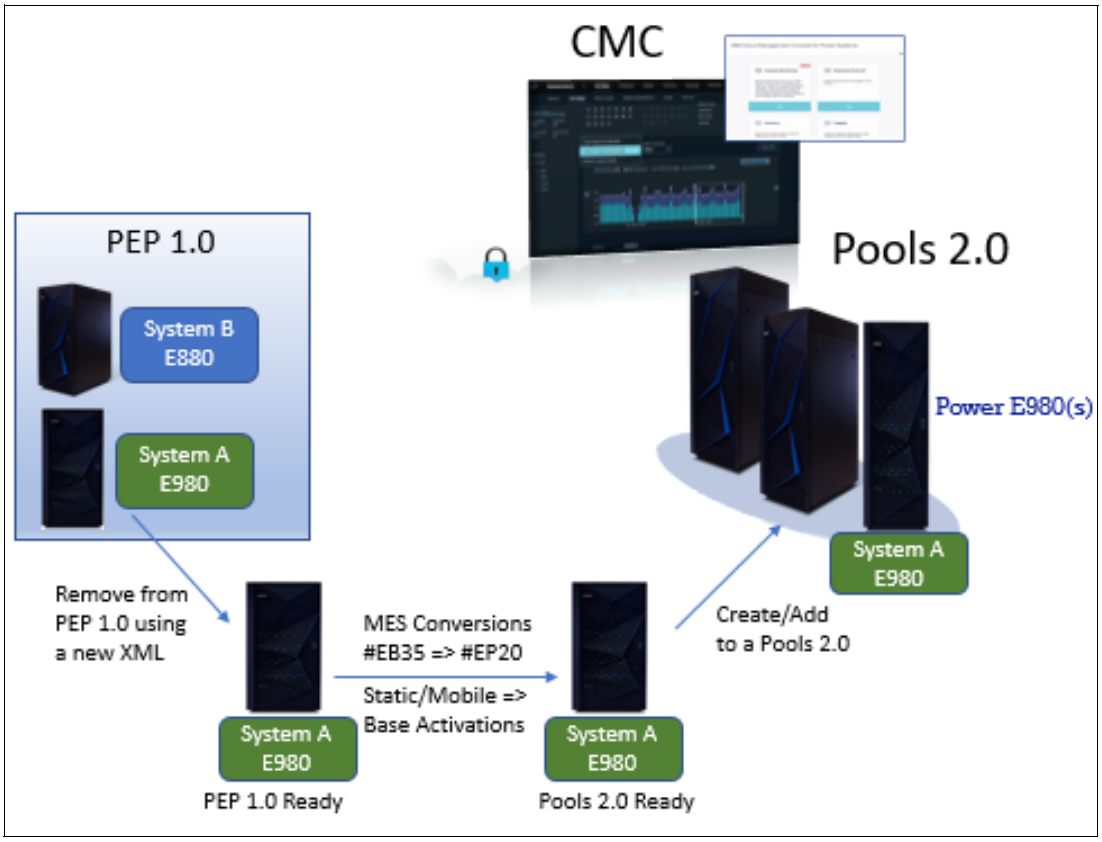


Figure 2-22 Migrating from PEP 1.0 to PEP 2.0

Before a server can be removed from a PEP 1.0 pool, all assigned mobile activations must be removed from that server and returned to the pool. When a system is removed from a PEP 1.0 pool, customers must complete the PEP Change Request Form and send a copy to the Power Systems CoD Project Office (`pcod@us.ibm.com`). When this request is processed, a new XML file (for the PEP 1.0 pool) is generated. When the new XML file is ready, it must be downloaded and applied by using the PEP 1.0 pool's controller HMC with the Update Pool operation. The controller HMC clears the PEP 1.0 pool configuration from systems that are specified in the XML file.

For more information about PEP 2.0, see the following resources:

- *IBM Power Systems Private Cloud with Shared Utility Capacity: Featuring Power Enterprise Pools 2.0*, SG24-8478
- Get Started with IBM Cloud Management Console for Power Systems, found at:

  `https://ibmcmc.zendesk.com/hc/en-us/articles/235776268-Get-Started-with-IBM-Cloud-Management-Console-for-Power-Systems`

3

Planning for IBM PowerVM

This chapter includes guidance for planning for PowerVM before you start to implement the solution.

This chapter covers the following topics:

- PowerVM prerequisites
- Processor virtualization planning
- Memory virtualization planning
- Virtual I/O Server planning
- Storage virtualization planning
- Network virtualization planning
- Further considerations

3.1 PowerVM prerequisites

PowerVM requires a valid license (feature code) before its features can be used. All IBM Power9 and IBM Power10 processor-based systems have PowerVM Enterprise Edition features.

The following sections present the hardware and operating system requirements that are associated with available PowerVM features.

3.1.1 Hardware requirements

PowerVM features are supported on most of the Power offerings with a few exceptions.

The availability of Capacity on Demand (CoD) offerings varies based on the Power server model. Support for CoD features can be found in 2.9, “Capacity on Demand” on page 66 and 2.10, “Power Enterprise Pools” on page 70.

Support for a few PowerVM features is discontinued. Table 1-3 on page 6 lists PowerVM features that are discontinued.

Hardware Management Console (HMC) hardware and supported code combinations for Power servers can be found at Power Code Matrix - Supported HMC Hardware, found at:

https://www.ibm.com/support/pages/node/6554904

3.1.2 Software requirements

PowerVM supports running AIX, IBM i, and Linux operating systems on Power servers. PowerVM offers Virtual I/O Server (VIOS) to facilitate I/O virtualization for client VMs. The supported versions of the operating systems, system firmware, I/O adapter firmware, HMC code level, and VIOS code levels depend on the Power server model.

The supported OS versions in PowerVM are as follows:

- AIX

  AIX 7.1, AIX 7.2, and AIX 7.3 and later
- IBM i

  IBM i 7.2, IBM i 7.3, IBM i 7.4, and IBM i 7.5 and later
- Linux
  - Red Hat Enterprise Linux V7 for Power, and RHEL 8 or later
  - SUSE Linux Enterprise Server 12, and SUSE Linux Enterprise Server 15 or later

Supported code combinations of HMC and system firmware levels for all IBM Power Systems are listed in the Power Code Matrix, found at:

https://esupport.ibm.com/customercare/flrt/mtm

For compatible system software combinations with POWER processors, see System Software Maps, found at:

https://www.ibm.com/support/pages/system-software-maps

Plan a successful Power system upgrade or migration by finding the minimum system software requirements at *IBM Power Systems Prerequisites*, found at:

https://esupport.ibm.com/customercare/iprt/home

3.2 Processor virtualization planning

PowerVM hypervisor (PHYP) can map a whole physical processor core or it can time slice a physical processor core. PHYP time slices shared processor partitions (also known as IBM Micro-Partitioning) on the physical CPUs by dispatching and undispatching the various virtual processors for the partitions that run in the shared pool. The minimum processing capacity per processor is 0.05 of a physical processor core, with a further granularity of 0.01. The PHYP uses a 10 millisecond (ms) time slicing dispatch window for scheduling all shared processor partitions' virtual processor queues to the PHYP physical processor core queues.

Partitions are created by using the HMC or PowerVM NovaLink and orchestrated by IBM Power Virtualization Center (PowerVC). When you start creating a partition, you must choose between a shared processor and a dedicated processor logical partition (LPAR).

3.2.1 Dedicated processors planning

Dedicated-processor LPARs can be allocated only in whole numbers. Therefore, the maximum number of dedicated-processor LPARs in a system is equal to the number of physical activated processors.

For dedicated processor partitions, you configure these attributes:

- Minimum number of processors
- Allocated number of processors
- Maximum number of processors

Allocated processors define the number of processors that you want for this partition. When the partition is activated, the hypervisor tries to allocate this number of processors to the partition. If not enough available processors are left in the system, then the hypervisor tries to allocate as many as possible from the remaining capacity. If the available processors in the system are less than the minimum value, the partition cannot be activated. Minimum and maximum values also set the limits of dynamic logical partitioning (DLPAR) operations while the partition is active. You cannot allocate more processors than the maximum value to the partition, and you can change only minimum and maximum values when the partition is inactive.

Consider setting the maximum value high enough to support the future requirements of growing workloads. Similarly, ensure that the minimum value is set low enough in case you must reduce the resources of this partition, but also set the minimum value high enough to prevent the scenario where the application is starving for CPU because the partition is activated with fewer processors than the number that is required for the application.

When a dedicated processor partition is powered off, its processors are donated to the default shared processor pool (SPP) by default. It is possible to disable this attribute in the partition properties windows. It also is possible to enable donating unused processing cycles of dedicated processors while the dedicated processor partition is running. You can change these settings at any time without having to shut down and restart the LPAR.

Processor resources within PEP 2.0 are tracked based on the assignment of dedicated processors to active partitions. If sharing of unused capacity is not enabled, the whole core is marked as consumed. If the processor is set to dedicated-donating mode, then actual consumption is reported like shared processor VMs. It is a best practice to allow sharing of unused capacity of dedicated processors for cost efficiency, especially in PEP 2.0 environments.

3.2.2 Shared processors planning

For shared processor partitions, you configure these additional attributes:

- Minimum, wanted, and maximum *processing units of capacity*
- The processing sharing mode, either *capped* or *uncapped*
- Minimum, wanted, and maximum *virtual processors*

Processing units of capacity

Processing capacity can be configured in fractions of 0.01 processors. The minimum amount of processing capacity that must be assigned to a micro-partition is 0.05 processors. On the HMC, processing capacity is specified in terms of processing units. The minimum capacity of 0.05 processors is specified as 0.05 processing units. To assign a processing capacity that represents 75% of a processor, 0.75 processing units are specified on the HMC.

On a system with two processors, a maximum of 2.0 processing units can be assigned to a micro-partition. Processing units that are specified on the HMC are used to quantify the minimum, wanted, and maximum amount of processing capacity for a shared processor partition.

After a shared processor partition is activated, processing capacity is usually referred to as capacity entitlement or entitled capacity. A shared processor partition is guaranteed to receive its capacity entitlement under all systems and processing circumstances.

Capacity entitlement must be correctly configured for normal production operation and, if capped, to cover workload during peak time. Having enough capacity entitlement is important to not impact operating system performance.

Capped and uncapped mode

Shared processor partitions have a specific processing mode that determines the maximum processing capacity that is given to them from their SPP.

The processing modes are as follows:

Uncapped mode The processing capacity can exceed the entitled capacity when extra resources are available in their SPP. Extra capacity is distributed on a weighted basis. An uncapped weight value is assigned to each uncapped partition when it is created.

Capped mode The processing capacity that is given can never exceed the entitled capacity of the shared processor partition.

If multiple uncapped partitions are competing for more processing capacity, the hypervisor distributes the remaining unused processor capacity in the processor pool to the eligible partitions in proportion to their uncapped weight. The higher the uncapped weight value, the more processing capacity the partition receives.

The uncapped weight must be an integer 0 - 255. The default uncapped weight for uncapped micro-partitions is 128. Uncapped weight provides information to the hypervisor on how unused capacity must be distributed across partitions. A partition with an uncapped weight of 100 is 100 times more likely to receive some of the unused capacity than a partition with an uncapped weight of 1.

Important: If you set the uncapped weight at 0, the hypervisor treats the micro-partition as a capped micro-partition. A micro-partition with an uncapped weight of 0 cannot be allocated more processing capacity beyond its entitled capacity.

3.2.3 Virtual processors planning

A *virtual processor* is a depiction or a representation of a physical processor that is presented to the operating system that runs in a micro-partition. The processing entitlement capacity that is assigned to a micro-partition, whether it is a whole or a fraction of a processing unit, is distributed by the server firmware equally between the virtual processors within the micro-partition to support the workload. For example, if a micro-partition has 1.60 processing units and two virtual processors, each virtual processor has the capacity of 0.80 processing units.

A virtual processor cannot have a greater processing capacity than a physical processor. The capacity of a virtual processor is equal to or less than the processing capacity of a physical processor.

A micro-partition must have enough virtual processors to satisfy its assigned processing capacity. This capacity can include its entitled capacity and any additional capacity beyond its entitlement if the micro-partition is uncapped.

So, the upper boundary of processing capacity in a micro-partition is determined by the number of virtual processors that it possesses. For example, if you have a partition with 0.50 processing units and one virtual processor, the partition cannot exceed 1.00 processing units. However, if the same partition with 0.50 processing units is assigned two virtual processors and processing resources are available, the partition can use an extra 1.50 processing units.

The maximum number of processing units that can be allocated to a virtual processor is always 1.00. Additionally, the number of processing units cannot exceed the total processing unit within an SPP.

Number of virtual processors

In general, the value of the minimum, wanted, and maximum virtual processor attributes must parallel the values of the minimum, wanted, and maximum capacity attributes in some fashion. A special allowance must be made for uncapped micro-partitions because they are allowed to consume more than their capacity entitlement.

If the micro-partition is uncapped, the administrator might want to define the wanted and maximum virtual processor attributes greater than the corresponding capacity entitlement attributes. The exact value is installation-specific, but 50 - 100 percent more is reasonable.

In general, it is a best practice to assign enough processing units to an uncapped partition to satisfy average workloads and set virtual processors high enough to address peak demands.

Because the number of virtual processors defines the number of physical cores that the partition has access to, it also sets the number of simultaneous multithreading (SMT) threads that are available to the partition. Therefore, you might want to adjust the number of virtual processors based on application requirements.

Selecting the optimal number of virtual processors depends on the workload in the partition. A high number of virtual processors might negatively affect the system performance. The number of virtual processors also can impact software licensing, for example, if the subcapacity licensing model is used.

Virtual processor folding

Virtual processor folding effectively puts idle virtual processors into a hibernation state so that they do not consume any resources. This feature provides several benefits, such as improved processor affinity, reduced hypervisor workload, and increased average time a virtual processor runs on a physical processor.

The characteristics of the virtual processor folding feature are:

- Idle virtual processors are not dynamically removed from the partition. They are *hibernated*, and only awoken when more work arrives.
- This feature provides no benefit when partitions are busy.
- If the feature is turned off, all virtual processors that are defined for the partition are dispatched to physical processors.
- Virtual processors that have attachments, such as **`bindprocessor`** or **`rset`** command attachments in AIX, are not excluded from being disabled.
- The feature can be turned off or on, and the default is on.

When a virtual processor is disabled, threads are not scheduled to run on it unless a thread is bound to that processor.

Virtual processor folding is controlled through the **`vpm_xvcpus`** tuning setting, which can be configured by using the **`schedo`** command.

3.2.4 Shared processor pools capacity planning

This section describes the capacity attributes of SPPs and provides examples of the capacity resolution according to the server load.

SPPs are described in 2.1.3, “Shared processors” on page 34 and 2.1.5, “Multiple shared processor pools” on page 36.

Capacity attributes

The following attributes are used to calculate the pool capacity of SPPs:

- Maximum Pool Capacity (MPC)

  Each SPP has a maximum capacity that is associated with it. The MPC defines the upper boundary of the processor capacity that can be used by the set of micro-partitions in the SPP. The MPC must be represented by a whole number of processor units.

- Reserved Pool Capacity (RPC)

  The system administrator can assign an entitled capacity to an SPP to reserve processor capacity from the physical SPP for the express usage of the micro-partitions in the SPP. The RPC is in addition to the processor capacity entitlements of the individual micro-partitions in the SPP. The RPC is distributed among uncapped micro-partitions in the SPP according to their uncapped weighting. The default value for the RPC is zero.

- Entitled Pool Capacity (EPC)

  The EPC of an SPP defines the guaranteed processor capacity that is available to the group of micro-partitions in the SPP. The EPC is the sum of the entitlement capacities of the micro-partitions in the SPP plus the RPC.

The default shared processor pool

The default SPP (SPP_0) is automatically activated by the system and is always present. Its MPC is set to the capacity of the physical SPP. For SPP_0, the RPC is always 0.

The default SPP has the same attributes as a user-defined SPP except that these attributes are not directly under the control of the system administrator; their values are fixed.

The maximum capacity of SPP_0 can change indirectly through system administrator action such as powering on a dedicated-processor partition or dynamically moving physical processors in or out of the physical SPP.

Levels of processor capacity resolution

Two levels of processor capacity resolution are implemented by the PHYP and multiple shared processor pools (MSPP):

$Level_0$ The first level, $Level_0$, is the resolution of capacity within the same SPP. Unused processor cycles from within an SPP are harvested and then redistributed to any eligible micro-partition within the same SPP.

$Level_1$ When all $Level_0$ capacities are resolved within the MSPP, the hypervisor harvests unused processor cycles and redistributes them to eligible micro-partitions regardless of the MSPP structure. $Level_1$ is the second level of processor capacity resolution.

Important: When user-defined SPPs are configured, the MPC is not deducted from default SPP (SPP_0). The default pool size stays the same. If the MPC is bigger than the sum of entitled capacity in the pool and RPC, partitions in the user-defined SPP might still compete for more processing capacity with partitions that are not in the same user-defined pool.

3.2.5 Software licensing in a virtualized environment

The following sections describe the factors to be considered when you plan the license model that you will use. A licensing factors summary is presented at the end.

Licensing factors in a virtualized system

With the mainstream adoption of virtualization, more independent software vendors (ISVs) are adapting their licensing to accommodate the new virtualization technologies. Several different models exist, varying with the ISVs. When you calculate the cost of licensing and evaluate which virtualization technology to use, consider the following factors:

- ISV recognition of virtualization technology and capacity capping method
- ISV subcapacity licensing available for selected software products
- ISV method for monitoring and management of subcapacity licensing
- ISV flexibility as license requirements change

Cost of software licenses

A careful consideration of the licensing factors in advance can help reduce the overall cost in providing business applications. Traditional software licensing is based on a fixed machine with a fixed number of resources. The new PowerVM technologies present some challenges to this model:

- It is possible to migrate partitions between different physical machines (with different speeds and numbers of total processors that are activated).
- Consider a number of partitions, which, at different times, are all using four processors. However, they can all be grouped by using multiple SPP technologies, which cap the overall CPU always at four CPUs in total.

When the ISV support for these technologies is in place, it is anticipated that it will be possible to increase the utilization within a fixed cost of software licenses.

Active processors and hardware boundaries

The upper boundary for licensing is always the quantity of active processors in the physical system (assigned and unassigned) because only active processors can be real engines for software.

Most software vendors consider each partition as a stand-alone server and depending on whether it is using dedicated processors or micro-partitioning, they license software per partition.

The quantity of processors for a certain partition can vary over time, for example, with dynamic partition operations. But, the overall licenses must equal or exceed the total number of processors that are used by the software at any point. If you are using uncapped micro-partitions, then the licensing must consider the fact that the partition can use extra processor cycles beyond the initial capacity entitlement.

Capacity capping

Two kinds of models for licensing software are available:

- A pre-pay license based on server capacity or number of users.
- A post-pay license based on auditing and accounting for actual capacity that is used.

Most software vendors offer the pre-pay method, and the question that they ask is about how much capacity a partition can use. The following sections illustrate how to calculate the amount of processing power that a partition can use.

Dedicated or dedicated-donating partitions

In a partition with dedicated processors, the initial licensing must be based on the number of processors that are assigned to the partition at activation. Depending on the partition profile maximums, if extra active processors or Capacity Upgrade on Demand (CUoD) processors are available in the system, these processors can be added dynamically, which allows operators to increase the quantity of processors.

Consider the number of software licenses before any additional processors are added, even temporarily, for example, with dynamic partition operations. Some ISVs might require licenses for the maximum number of processors for each of the partitions where the software is installed (the maximum quantity of processors in the partition profile).

Sharing idle processor cycles from running dedicated processor partitions does not change the licensing considerations.

Capacity capping of micro-partitions

Several factors must be considered when you calculate the capacity of micro-partitions. To allow the hypervisor to create micro-partitions, the physical processors are presented to the operating system as virtual processors. As micro-partitions are allocated processing time by the hypervisor, these virtual processors are dispatched on physical processors on a time-share basis.

With each logical processor mapping to a physical processor, the maximum capacity that an uncapped micro-partition can use is the number of available virtual processors, with the following assumptions:

- This capacity does not exceed the number of active processors in the physical system.
- This capacity does not exceed the available capacity in the SPP.

The following sections describe the different configurations that are possible and the licensing implications of each one.

Capped micro-partition

For a micro-partition, the wanted entitled capacity is a guaranteed capacity of computing power that a partition is given on activation. For a capped micro-partition, the entitled capacity also is the maximum processing power that the partition can use.

By using dynamic LPAR operations, you can vary the entitled capacity between the maximum and minimum values in the profile.

Uncapped micro-partition without MSPP technology

The entitled capacity that is given to an uncapped micro-partition is not necessarily a limit on the processing power. An uncapped micro-partition can use more than the entitled capacity if some resources within the system are available.

In this case, on a Power server that uses SPPs or that uses only the default SPP, the limiting factor for uncapped micro-partition is the number of virtual processors. The micro-partition can use up to the number of physical processors in the SPP because each virtual processor is dispatched to a physical processor.

With a single pool, the total resources that are available in the SPP are equal to the activated processors in the machine minus any dedicated (nondonating) partitions. The assumption is that at a point all other partitions are idle.

The total licensing liability for an uncapped partition without MSPP technology is either the number of virtual processors or the number of processors in the default SPP, whichever is smallest.

Uncapped micro-partition with MSPP technology

Similarly, the entitled capacity for an uncapped micro-partition is not necessarily a limit on the processing power. An uncapped micro-partition can use more than the entitled capacity if some resources are available within the system.

By using MSPP technology, it is possible to group micro-partitions and place a limit on the overall group maximum processing units. After an SPP group is defined, operators can group specific micro-partitions that are running the same software (if the software licensing terms permit it). This approach allows a pool of capacity that can be shared among several different micro-partitions.

System with CoD processors

Processors in the CoD pool do not count for licensing purposes until the following events happen:

- They become temporarily or permanently active and assigned to partitions.
- They become temporarily or permanently active in systems with PowerVM technology, and they can be used by micro-partitions.

Clients can provision licenses of selected software for temporary or permanent usage on their systems. Such licenses can be used to align with the possible temporary or permanent usage of CoD processors in existing or new AIX, IBM i, or Linux partitions.

Summary of licensing factors

Depending on the licensing model that is supported by the software vendor, it is possible to work out licensing costs based on these factors:

- Capped versus uncapped micro-partitions.
- Number of virtual processors.
- Unused processing cycles that are available in the machine, from dedicated-donating partitions and other micro-partitions.
- Multiple shared processor pool maximum.
- Active physical processors in the system.

An example of the license boundaries is illustrated in Figure 3-1 on page 87.

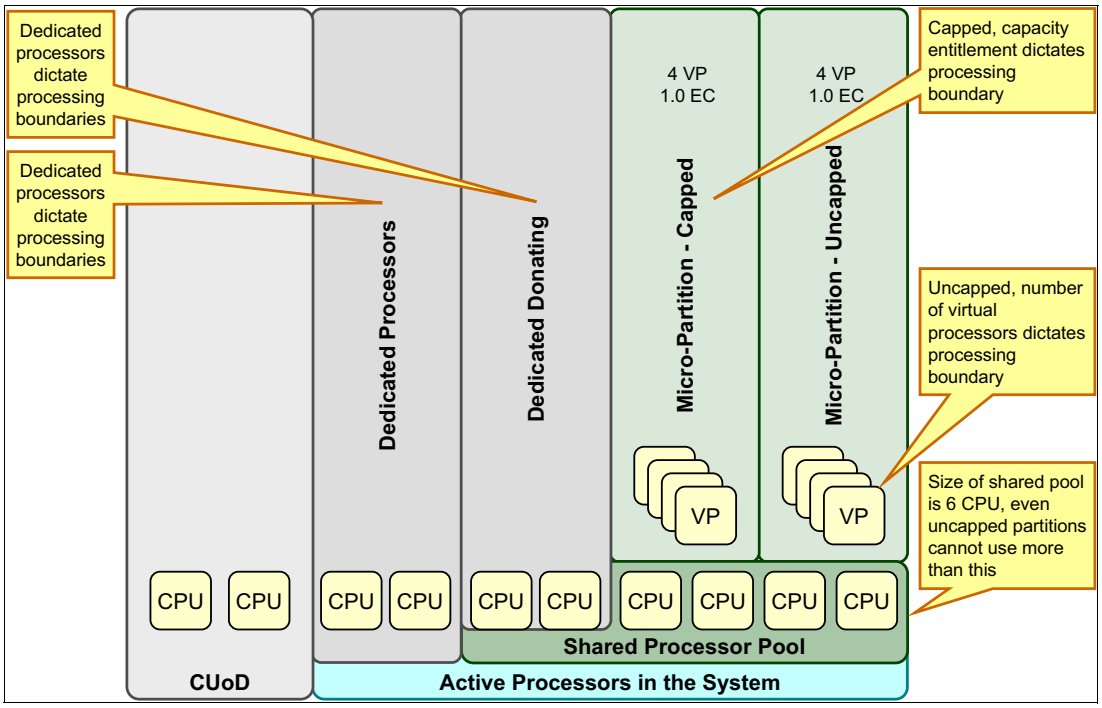


Figure 3-1 License boundaries with different processor and pool modes

IBM i software licensing

It is possible to use workload groups to limit the processing capacity of a workload to a subset of processor cores in a partition. This capability requires the workload groups' PTFs.

Therefore, workload groups can be used to reduce license costs for a processor usage type-licensed program by completing the following steps:

- Create a workload group with a maximum processor core limit that is less than the number of processor cores that are configured for the partition.
- Add the licensed program to the newly created workload group.
- Identify the workloads that are associated with the licensed program and associate the workloads with the newly created workload group.

The licensed program owner must accept the reduced processor core capacity.

Linux software licensing

The license terms and conditions of Linux operating system distributions are provided by the Linux distributor, but all base Linux operating systems are licensed under the GPL. Distributor pricing for Linux includes media, packaging, shipping, and documentation costs, and they can offer extra programs under other licenses, and bundled service and support.

Clients or authorized IBM Business Partners are responsible for the installation of the Linux operating system, with orders handled according to license agreements between the client and the Linux distributor.

Clients must consider the quantity of virtual processors in micro-partitions for scalability and licensing purposes (uncapped partitions) when Linux is installed in a virtualized Power server.

Each Linux distributor sets its own pricing method for their distribution, service, and support. For more information, check the distributor's website and the following resources:

- SUSE Linux Enterprise Server, found at:

  https://www.suse.com/products/server/

- Red Hat, found at:

  https://www.redhat.com/en

For more information about Linux licensing, contact an IBM sales representative and see Enterprise Linux on Power, found at:

https://www.ibm.com/it-infrastructure/power/os/linux

3.3 Memory virtualization planning

This section describes the points that you need to plan and verify before you configure the server and implement the Active Memory Expansion (AME) memory virtualization features in your environment.

3.3.1 Hypervisor memory planning

The PHYP uses some of the memory that is activated in a Power server to manage memory that is assigned to individual partitions, manage I/O requests, and support virtualization requests. The amount of memory that is required by the hypervisor to support these features varies based on various configuration options that are chosen.

The assignment of the memory to the hypervisor ensures secure isolation between LPARs because the only allowed access to the memory contents is through security-validated hypervisor interfaces. In Figure 3-2, 128 GB is installed in the system, 128 GB is licensed memory (Configurable), and 3.5 GB (Reserved) memory is assigned to the hypervisor.

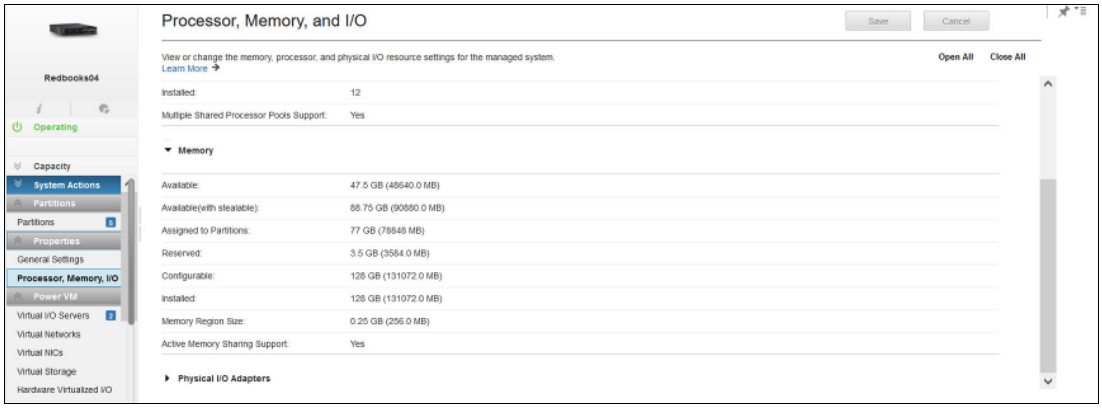


Figure 3-2 System memory properties

Components that contribute to hypervisor memory usage

The three main components that contribute to the overall usage of memory by the hypervisor are:

1. Memory that is required for hardware page tables (HPTs).
2. Memory that is required to support I/O devices.
3. Memory that is required for virtualization.

Memory usage for hardware page table

Each partition on the system has its own HPT that contributes to hypervisor memory usage. The HPT is used by the operating system to translate from effective addresses to physical real addresses in the hardware. This translation from effective to real addresses allows multiple operating systems to all run simultaneously in their own logical address space. The amount of memory for the HPT is based on the maximum memory size of the partition and the HPT ratio. The default HPT ratio is either 1/64 of the maximum (for IBM i partitions) or 1/128 (for AIX, VIOS, and Linux partitions) of the maximum memory size of the partition. AIX, VIOS, and Linux use larger page sizes (16 K, 64 K, and such) instead of using 4 K pages. Using larger page sizes reduces the overall number of pages that must be tracked so the overall size of the HPT can be reduced. For example, for an AIX partition with a maximum memory size of 256 GB, the HPT is 2 GB.

When a partition is defined, the maximum memory size that is specified must be based on the amount of memory that can be dynamically added to the partition (DLPAR) without having to change the configuration and restart the partition.

Memory usage for I/O devices

In support I/O operations, the hypervisor maintains structures that are called the Translation Control Entries (TCEs), which provide an information path between I/O devices and partitions. The TCEs provide the address of the I/O buffer, indication of read versus write requests, and other I/O-related attributes. Many TCEs per I/O device are in use, so multiple requests can be active simultaneous to the same physical device. For physical I/O devices, the base amount of space for the TCEs is defined by the hypervisor, based on the number of I/O devices that are supported.

Memory usage for virtualization features

Virtualization requires extra memory to be allocated by the hypervisor for hardware statesave areas and all the various virtualization technologies. For example, on Power8 processor-based servers and later servers, each processor core supports up to eight SMT threads of execution, and each thread contains over 80 different registers. The hypervisor must set aside save areas for the register contents for the maximum number of virtual processors that are configured. The greater the number of physical hardware devices, the greater the number of virtual devices, the greater the amount of virtualization, and the more hypervisor memory is required. For efficient memory consumption, wanted and maximums for various attributes (processors, memory, and virtual adapters) must be based on business needs, and not set to values that are higher than actual requirements.

Predicting memory usage

The IBM System Planning Tool (SPT) is a resource that can be used to estimate the amount of hypervisor memory that is required for a specific server configuration. After the SPT executable file is downloaded and installed, a configuration can be defined by selecting the appropriate hardware platform, installed processors, and memory, which define partitions and partition attributes. Given a configuration, the SPT can estimate the amount of memory that will be assigned to the hypervisor. This capability can help to change an existing configuration or when new servers are deployed.

For more information about SPT, see IBM System Planning Tool for Power processor-based systems, found at:

https://www.ibm.com/support/pages/ibm-system-planning-tool-power-processor-based-systems-0

3.3.2 Active Memory Expansion planning

AME is described in 2.2.2, “Active Memory Expansion” on page 39.

When a partition with AME is configured, the following two settings define how much memory is available:

Physical memory The amount of physical memory that is available to the partition. Usually, it corresponds to the wanted memory in the partition profile.

Memory expansion factor Defines how much of the physical memory is expanded.

Tip: The memory expansion factor can be defined individually for each partition.

AME relies on compression of in-memory data to increase the amount of data that can be placed into memory and thus expands the effective memory capacity of Power servers. The in-memory data compression is managed by the operating system, and this compression is transparent to applications and users.

The amount of memory that is available to the operating system can be calculated by multiplying the physical memory with the memory expansion factor. For example, in a partition that has 10 GB of physical memory and configured with a memory expansion factor of 1.5, the operating system sees 15 GB of available memory.

The compression and decompression activities require CPU cycles. Therefore, when AME is enabled, spare CPU resources must be available in the partition for AME.

AME does not compress file cache pages and pinned memory pages.

If the expansion factor is too high, the target-expanded memory size cannot be achieved and a memory deficit forms. The effect of a memory deficit is the same as the effect of configuring a partition with too little memory. When a memory deficit occurs, the operating system might have to resort to paging out virtual memory to the paging space.

Note: When AME is enabled, by default the AIX operating system uses 4 KB pages. However, if you are running IBM AIX 7.2 with Technology Level 1 or later on a Power9 or a Power10 processor-based server, you can use the **vmo** command with the **ame_mpsize_support** parameter to enable 64 KB page size.

AME factor

You can configure the degree of memory expansion that you want to achieve for the LPAR by setting the AME factor in a partition profile of the LPAR. The expansion factor is a multiplier of the amount of memory that is assigned to the LPAR.

When AME is configured, a single configuration option must be set for the LPAR, which is the memory expansion factor. An LPAR's memory expansion factor specifies the target effective memory capacity for the LPAR. This target memory capacity provides an indication to the operating system of how much memory is made available with memory compression. The target memory capacity that is specified is referred to as the expanded memory size. The memory expansion factor is specified as a multiplier of an LPAR's true memory size, as shown in the following equation:

```
LPAR_expanded_mem_size = LPAR_true_mem_size * LPAR_mem_exp_factor
```

For example, an LPAR's memory expansion factor of 2.0 indicates that memory compression must be used to double the LPAR's memory capacity. If an LPAR is configured with a memory expansion factor of 2.0 and a memory size of 20 GB, then the expanded memory size for the LPAR is 40 GB, as shown in the following equation:

```
40 GB = 20 GB * 2.0
```

The operating system compresses enough in-memory data to fit 40 GB of data into 20 GB of memory. The memory expansion factor and the expanded memory size can be dynamically changed at run time by using the HMC through dynamic LPAR operations. The expanded memory size is always rounded down to the nearest logical memory block (LMB) multiple.

Note: You do not need to check whether your application is certified for AME; it is hidden within the AIX kernel.

Memory deficit

When the memory expansion factor for an LPAR is configured, it is possible that the chosen memory expansion factor is too large and cannot be achieved based on the compressibility of the workload.

When the memory expansion factor for an LPAR is too large, then a memory expansion deficit forms, which indicates that the LPAR cannot achieve its memory expansion factor target. For example, if an LPAR is configured with a memory size of 20 GB and a memory expansion factor of 1.5, it results in a total target-expanded memory size of 30 GB. However, the workload that runs in the LPAR does not compress well, and the workload's data compresses only by a ratio of 1.4 to 1. In this case, it is impossible for the workload to achieve the targeted memory expansion factor of 1.5. The operating system limits the amount of physical memory that can be used in a compressed pool up to a maximum of 95%. This value can be adjusted by using the **vmo** command with the **ame_min_ucpool_size** parameter. In this example with the LPAR memory size as 20 GB, if the **ame_min_ucpool_size** parameter value is set to 90, 18 GB are reserved for compressed pool. The maximum achievable expanded memory size is 27.2 GB (2 GB + 1.4 x 18 GB). The result is a 2.8 GB shortfall. This shortfall is referred to as the *memory deficit*.

The effect of a memory deficit is the same as the effect of configuring an LPAR with too little memory. When a memory deficit occurs, the operating system cannot achieve the expanded memory target that is configured for the LPAR. In this case, the operating system might have to resort to paging out virtual memory pages to paging space. Thus, in the previous example, if the workload uses more than 27.2 GB of memory, the operating system starts paging out virtual memory pages to paging space.

To get an indication of whether a workload can achieve its expanded memory size, the operating system reports a memory deficit metric. This deficit is a "hole" in the expanded memory size that cannot be achieved. If this deficit is zero, the target memory expansion factor can be achieved, and the LPAR's memory expansion factor is configured correctly. If the expanded memory deficit metric is nonzero, then the workload falls short of achieving its expanded memory size by the size of the deficit.

To eliminate a memory deficit, the LPAR's memory expansion factor must be reduced. However, reducing the memory expansion factor reduces the LPAR's expanded memory size. Thus, to keep the LPAR's expanded memory size the same, the memory expansion factor must be reduced and more memory must be added to the LPAR. Both the LPAR's memory size and memory expansion factor can be changed dynamically.

AME planning

The benefit of AME to a workload varies based on the workload's characteristics. Some workloads can get a higher level of memory expansion than other workloads. The AME Planning and Advisory Tool **amepat** helps plan the deployment of a workload in the AME environment. It also provides guidance on the level of memory expansion that a workload can achieve.

The AME Planning Tool (located in `/usr/bin/amepat`) serves two primary purposes. They are:

- To plan an initial AME configuration.
- To monitor and fine-tune an active AME configuration.

The AME Planning Tool can run on LPARs with and without AME enabled. In an LPAR where AME was not enabled, run **amepat** with a representative workload. Set **amepat** to monitor the workload for a meaningful period. For example, the **amepat** tool is set to run during a workload's peak resource usage. After it completes, the tool displays a report with various potential memory expansion factors and the expected CPU utilization attributable to an AME for each factor. The tool also provides a recommended memory expansion factor that seeks to maximize memory savings while minimizing extra CPU utilization.

Figure 3-3 on page 93 shows an **amepat** output sample report.

```
# amepat 5 2

Command Invoked                : amepat 2 5

Date/Time of invocation        : Wed Dec  2 11:29:29 PAKST 2009
Total Monitored time           : 10 mins 58 secs
Total Samples Collected        : 5

System Configuration:
---------------------
Partition Name                 : aixfvt19
Processor Implementation Mode  : POWER5
Number Of Logical CPUs         : 8
Processor Entitled Capacity    : 4.00
Processor Max. Capacity        : 4.00
True Memory                    : 4.25 GB
SMT Threads                    : 2
Shared Processor Mode          : Disabled
Active Memory Sharing          : Disabled
Active Memory Expansion        : Disabled

System Resource Statistics:    Average       Min         Max
---------------------------    -----------   ----------  -----------
CPU Util (Phys. Processors)    2.00 [ 50%]   1.00 [ 25%] 3.00 [ 75%]
Virtual Memory Size (MB)       1366 [ 31%]   1113 [ 26%] 2377 [ 55%]
True Memory In-Use (MB)        1758 [ 40%]   1234 [ 28%] 3834 [ 88%]
Pinned Memory (MB)             673 [ 15%]     673 [ 15%]  675 [ 16%]
File Cache Size (MB)           391 [  9%]     124 [  3%] 1437 [ 33%]
Available Memory (MB)          841 [ 65%]    1812 [ 42%] 3099 [ 71%]

Active Memory Expansion Modeled Statistics
------------------------------------------
Modeled Expanded Memory Size   :   4.25 GB
Average Compression Ratio      :   5.29

Expansion    Modeled True      Modeled              CPU Usage
Factor       Memory Size       Memory Gain          Estimate
---------    -------------     ------------------   -----------
     1.00          4.25 GB         0.00 KB [  0%]   0.00 [  0%]
     1.31          3.25 GB         1.00 GB [ 31%]   0.34 [  8%]
     1.55          2.75 GB         1.50 GB [ 55%]   0.39 [ 10%]
     1.89          2.25 GB         2.00 GB [ 89%]   0.45 [ 11%]
     2.12          2.00 GB         2.25 GB [112%]   0.50 [ 12%]
     2.43          1.75 GB         2.50 GB [143%]   0.65 [ 16%]
     2.83          1.50 GB         2.75 GB [183%]   0.70 [ 18%]

Active Memory Expansion Recommendation:
---------------------------------------
The recommended AME configuration for this workload is to configure the LPAR
with a memory size of 1.50 GB and to configure a memory expansion factor
of 2.83.  This will result in a memory gain of 183%. With this
configuration, the estimated CPU usage due to AME is approximately 0.50
physical processors, and the estimated overall peak CPU resource required for
the LPAR is 3.50 physical processors.

NOTE: amepat's recommendations are based on the workload's utilization level
during the monitored period. If there is a change in the workload's utilization
level or a change in workload itself, amepat should be run again.

The modeled Active Memory Expansion CPU usage reported by amepat is just an
estimate.  The actual CPU usage used for Active Memory Expansion may be lower
or higher depending on the workload.
```

Figure 3-3 An amepat output sample report

The report and recommendation can be a useful initial configuration for an AME deployment. In an LPAR where AME is enabled, **amepat** serves a similar purpose. When it is run at peak time for a representative workload, the tool provides a report with the actual CPU utilization attributable to AME at the current memory expansion factor. It also displays memory deficit information if it is present. Because the AME is enabled, the tool can also provide a more accurate representation of what CPU utilization levels can be expected at different memory expansion factors. A new recommendation based on this information is presented to the user.

For more information about the **amepat** report, see Active Memory Expansion (AME), found at:

https://www.ibm.com/docs/en/aix/7.3?topic=management-active-memory-expansion-ame

3.4 Virtual I/O Server planning

This section describes the details to consider for planning a VIOS.

3.4.1 Specifications that are required to create the VIOS

To activate the VIOS, the PowerVM Editions hardware feature is required. An LPAR with enough resources to share with other LPARs also is required. Table 3-1 shows a list of minimum hardware requirements that must be available to create the VIOS.

Table 3-1 Resources that are required for VIOS

Resource	Requirement
HMC	The HMC is required to create the LPAR and assign resources.
Storage adapter	The server LPAR needs at least one storage adapter.
Physical disk	The disk must be at least 30 GB. This disk can be shared.
Ethernet adapter	To route network traffic from Virtual Ethernet Adapters (VEAs) to a Shared Ethernet Adapter (SEA), you need an Ethernet adapter.
Memory	A general rule for the minimum memory requirement for VIOS 3.1 is 4 GB. A minimum current memory requirement might support a configuration with a minimum number of devices or a small maximum memory configuration. However, to support shared storage pools (SSPs), the minimum memory requirement is 4 GB. More devices increase the minimum current memory requirement.
Processor	At least 0.05 processing units are required.

Table 3-2 defines the limitations for storage management:

Table 3-2 Limitations for storage management

Category	Limit
Volume groups	4096 per system.
Physical volumes	1024 per volume group.
Physical disk	The disk must be at least 30 GB. This disk can be shared.

Category	Limit
Physical partitions	1024 per volume group.
Logical volumes	1024 per volume group.
LPARs	No limit.

Limitations and restrictions of the VIOS configuration

Consider the following items when you implement virtual SCSI (vSCSI):

- vSCSI supports the following connection standards for backing devices: Fibre Channel (FC), SCSI, SCSI RAID, iSCSI, SAS, SATA, Universal Serial Bus (USB), and IDE.
- The SCSI protocol defines mandatory and optional commands. Although vSCSI supports all the mandatory commands, not all the optional commands are supported.
- There might be utilization implications when you use vSCSI devices. Because the client/server model is made up of layers of function, vSCSI can consume more processor cycles when processing I/O requests.
- The VIOS is a dedicated LPAR that is used only for VIOS operations. Other applications cannot run in the VIOS LPAR.
- If there is a resource shortage, performance degradation might occur. If a VIOS is serving many resources to other LPARs, ensure that enough processor power is available. In case of high workload across VEAs and virtual disks, LPARs might experience delays in accessing resources.
- Logical volumes and files that are exported as vSCSI disks are always configured as single path devices on the client LPAR.
- Logical volumes or files that are exported as vSCSI disks that are part of the root volume group (rootvg) are not persistent if you reinstall the VIOS. However, they are persistent if you update the VIOS to a new Service Pack (SP). Therefore, before you reinstall the VIOS, ensure that you back up the corresponding clients' virtual disks. When exporting logical volumes, it is best to export logical volumes from a volume group other than the root volume group. When exporting files, it is best to create file storage pools and the virtual media repository in a parent storage pool other than the root volume group.

Consider the following items when you implement virtual adapters:

- Only Ethernet adapters can be shared.
- IP forwarding is not supported on the VIOS.
- The maximum number of virtual adapters can be any value 2 - 65,536. However, if you set the maximum number of virtual adapters to a high value, the server firmware requires more system memory to manage the virtual adapters. Setting the maximum number of virtual adapters to an excessive number might even lead an LPAR failing to activate.
- Consider the following items when you increase the virtual I/O slot limit:
  - The maximum number of virtual I/O slots that is supported on AIX, IBM i, and Linux partitions is 32,767.
  - The maximum number of virtual adapters can be any value 2 - 32767. However, higher maximum values require more system memory to manage the virtual adapters.

Sizing of processor and memory

The sizing of the processor and memory resources for VIOS depends on the amount and type of workload that the VIOS must process. For example, network traffic that goes through a SEA requires more processor resources than vSCSI traffic. Also, when a VIOS is used as a mover service partition (MSP), it requires more processor and memory resources during an active Live Partition Mobility (LPM) operation.

Table 3-3 can be used as a starting point for the environment.

Rules: The following examples are only starting points when you set up an environment that uses the VIOS for the first time. The actual sizing might vary depending on the level of the virtualization and configuration of the system.

Table 3-3 Virtual I/O Server sizing examples

Environment	CPU (example)	Virtual CPU (example)	Memory (example)
Small environment	0.25 - 0.5 processors (uncapped)	1 - 2	4 GB
Large environment	1 - 2 processors (uncapped)	2 - 4	6 GB
Environment that uses SSPs	At least one processor (uncapped)	4 - 6	8 GB

Monitoring: When the environment is in production, the processor and memory resources on the VIOS must be monitored regularly, and adjusted if necessary to make sure the configuration fits with workload. For more information about monitoring CPU and memory on the VIOS, see *IBM PowerVM Virtualization Managing and Monitoring*, SG24-7590.

The VIOS is designed for selected configurations that include specific models of IBM and other vendor storage products. Consult your IBM representative or IBM Business Partner for the latest information and included configurations.

List of supported adapters

Virtual devices that are exported to client partitions by the VIOS must be attached through supported adapters. An updated list of supported adapters and storage devices is available at the following websites:

- Adapter information by feature code for the 9043-MRX, 9080-HEX, 9105-22A, 9105-22B, 9105-41B, 9105-42A, 9786-22H, or 9786-42H system and EMX0 PCIe3 expansion drawers, found at:

  https://www.ibm.com/docs/en/power10/9080-HEX?topic=adapters-adapter-information-by-feature-code

- Adapter information by feature code for the 5105-22E, 9008-22L, 9009-22A, 9009-22G, 9009-41A, 9009-41G, 9009-42A, 9009-42G, 9040-MR9, 9080-M9S, 9223-22H, 9223-22S, 9223-42H, 9223-42S system, and EMX0 PCIe3 expansion drawers, found at:

  https://www.ibm.com/docs/en/power9/9080-M9S?topic=adapters-adapter-information-by-feature-code

Plan carefully before you begin the configuration and installation of your VIOS and client partitions. Depending on the type of workload and the needs of an application, it is possible to mix virtual and physical devices in the client partitions.

For more information about planning for the VIOS, see Planning for the Virtual I/O Server, found at:

https://www.ibm.com/docs/en/power10/9080-HEX?topic=server-planning

3.4.2 Redundancy considerations

This section describes requirements for providing high availability (HA) for VIOSs.

Redundancy options are available at several levels in the virtual I/O environment. Multipathing, mirroring, and RAID redundancy options exist for the VIOS and client LPARs. Ethernet link aggregation (LA) (also called Etherchannel) is also an option for the client LPARs, and the VIOS provides SEA failover and single-root I/O virtualization (SR-IOV) with virtual Network Interface Controllers (vNICs) failover. SR-IOV with vNIC is described in 2.4.4, “SR-IOV with virtual Network Interface Controller” on page 50.

Support for node failover (by using IBM PowerHA SystemMirror or VM Recovery Manager (VMRM)) is available for nodes that use virtual I/O resources.

This section contains information about redundancy for both the client LPARs and the VIOS. Although these configurations help protect the LPARs and VIOS from the failure of one of the physical components, such as a disk or network adapter, they might cause the client LPAR to lose access to its devices if the VIOS fails. The VIOS can be made redundant by running a second instance in another LPAR. When you run two instances of the VIOS, you can use logical volume mirroring (LVM), multipath input/output (MPIO), Network Interface Backup (NIB), or multipath routing with Dead Gateway Detection (DGD) in the client LPAR to provide HA access to virtual resources that are hosted in separate VIOS LPARs.

In a dual-VIOS configuration, vSCSI, virtual Fibre Channel (VFC) (NPIV), SEA, and SR-IOV with vNIC failover can be configured in a redundant fashion. This approach allows system maintenance such as restarts, software updates, or even reinstallation to be performed on a VIOS without causing outage to virtual I/O clients. This reason is the main one to implement dual VIOSs.

With proper planning and architecture implementation, maintenance can be performed on a VIOS and any external device to which it connects, such as a network or storage area network (SAN) switch, removing the layer of physical resource dependency.

When the client partition uses multipathing and SEA or SR-IOV with vNIC failover, no actions need to be performed on the client partition during the VIOS maintenance, or after it completes. This approach results in improved uptime and reduced system administration efforts for the client partitions.

Tip: A combination of multipathing for disk redundancy and SEA failover or SR-IOV with vNIC failover for network redundancy are industry best practices.

Upgrading and rebooting a VIOS, network switch, or SAN switch is simpler and more compartmentalized because the client no longer depends on the availability of all the environment.

In Figure 3-4, a client partition has vSCSI devices and a VEA that is backed by two VIOSs. The client has multipathing implemented across the vSCSI devices and SEA failover for the virtual Ethernet.

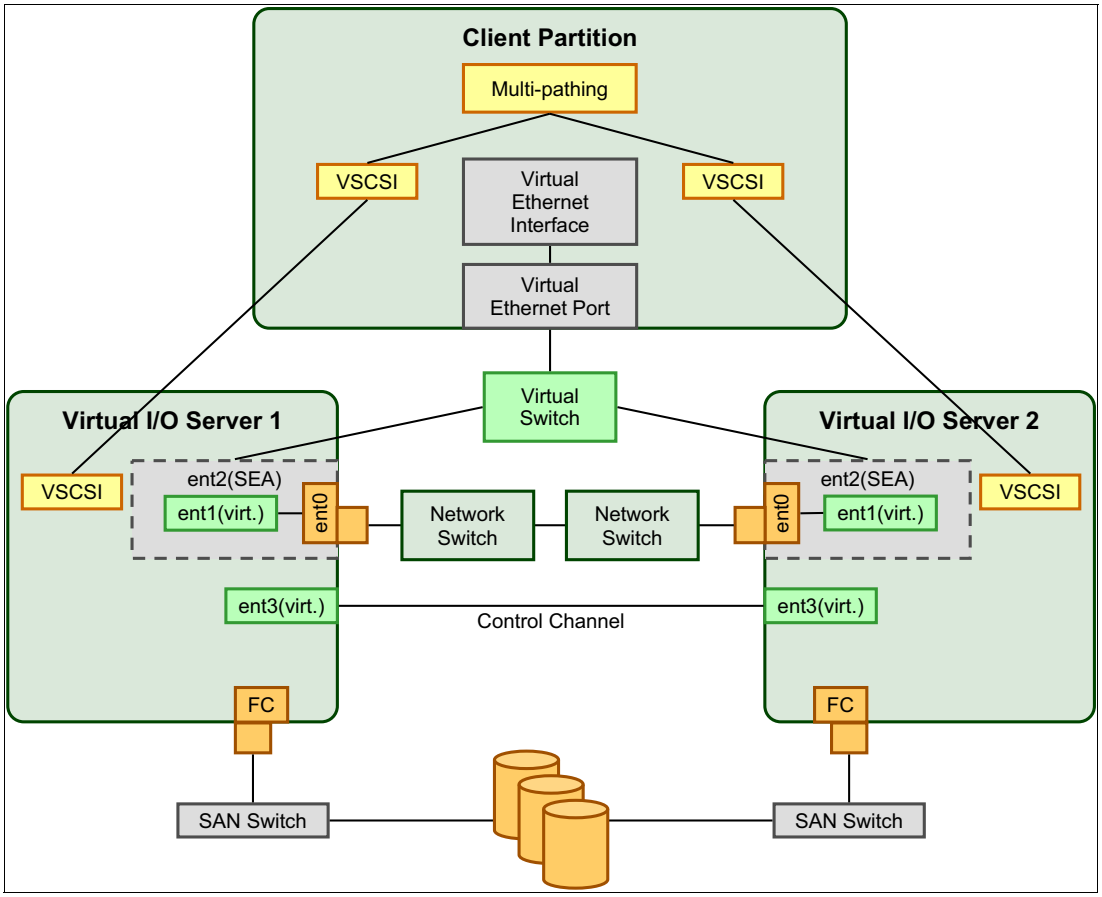


Figure 3-4 Redundant Virtual I/O Servers before maintenance

When VIOS 2 is shut down for maintenance, as shown in Figure 3-5 on page 99, the client partition continues to access the network and SAN storage through VIOS 1.

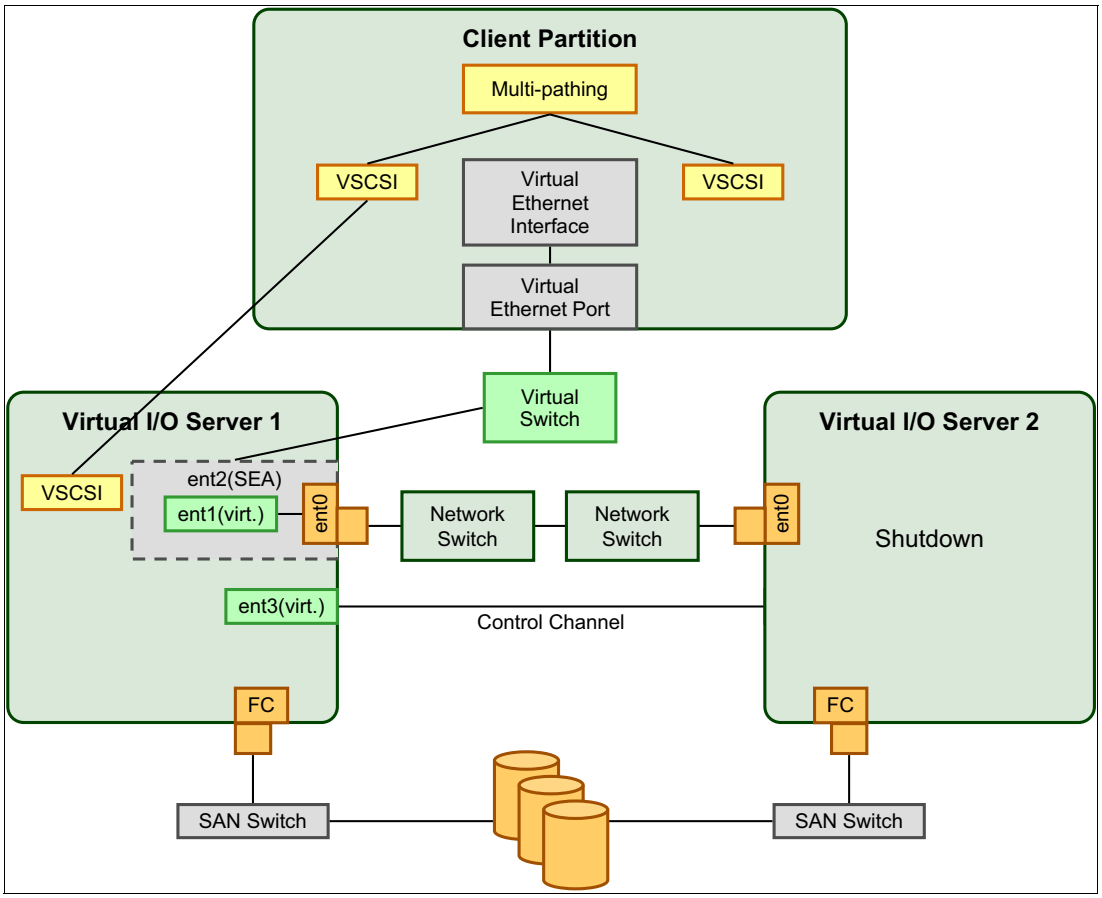


Figure 3-5 Redundant Virtual I/O Servers during maintenance

When VIOS 2 returns to a full running state, these events occur:

- An AIX client continues to use the MPIO path through VIOS 1 unless the MPIO path is manually changed to VIOS 2.
- An IBM i or Linux multipathing client, which uses a round-robin multipathing algorithm, automatically starts to use both paths when the path to VIOS 2 becomes operational again.
- If VIOS 2 is the primary SEA, client network traffic that goes through the backup SEA on VIOS 1 automatically resumes on VIOS 2.

In addition to continuous availability, a dual VIOS setup also separates or balances the virtual I/O load, which results in resource consumption across the VIOSs.

Virtual Ethernet traffic is generally heavier on the VIOS than vSCSI traffic. Virtual Ethernet connections generally take up more CPU cycles than connections through physical Ethernet adapters. The reason is that modern physical Ethernet adapters contain many functions to offload some work from the system's CPUs, for example, checksum computation and verification, interrupt modulation, and packet reassembly.

In a configuration that runs MPIO and a single SEA per VIOS, the traffic is typically separated so that the virtual Ethernet traffic goes through one VIOS and the vSCSI traffic goes through the other. This separation is done by defining the SEA trunk priority and the MPIO path priority.

Important: Do not turn off SEA threading on VIOS that might be used both for storage and network virtualization.

In an MPIO configuration with several SEAs per VIOS, you typically balance the network and vSCSI traffic between VIOSs.

Paths: Use the storage configuration commands to check that the preferred paths on the storage subsystem are in accordance with the path priorities that are set in the virtual I/O clients.

Figure 3-6 shows an example configuration where network and disk traffic are separated:

- VIOS 1 has priority 1 for the network and priority 2 for the disk.
- VIOS 2 has priority 2 for the network and priority 1 for the disk.

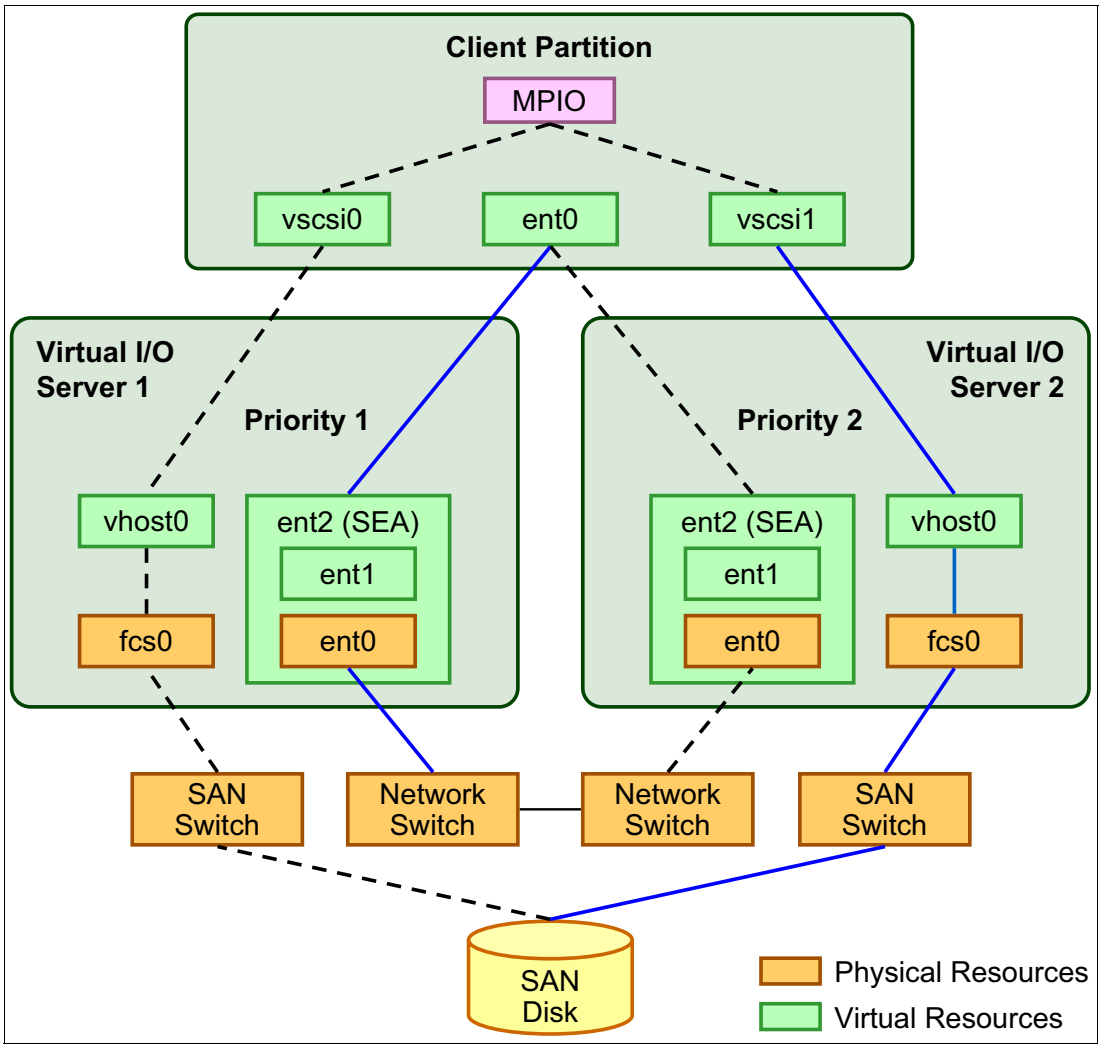


Figure 3-6 Separating disk and network traffic

3.5 Storage virtualization planning

The following sections explain how to plan storage virtualization in a PowerVM environment that uses vSCSI and VFC (NPIV).

Note: The configurations that are described in this section are not a complete list of all available supported configurations.

3.5.1 Virtual SCSI planning

By using vSCSI, client LPARs can share disk storage and tape or optical devices that are assigned to the VIOS LPAR.

Physical storage devices such as disk, tape, USB mass storage, or optical devices that are attached to the VIOS LPAR can be shared by one or more client LPARs. The VIOS provides access to storage subsystems by using logical unit numbers (LUNs) that are compliant with the SCSI protocol. The VIOS can export a pool of heterogeneous physical storage as a homogeneous pool of block storage in the form of SCSI disks. The VIOS is a storage subsystem. Unlike typical storage subsystems that are physically in the SAN, the SCSI devices that are exported by the VIOS are limited to the domain within the server. Therefore, although the SCSI LUNs are SCSI-compliant, they might not meet the needs of all applications, particularly those applications that exist in a distributed environment.

The following SCSI peripheral device types are supported:

- Disk that is backed by a logical volume.
- Disk that is backed by a file.
- Disk that is backed by a logical unit (LU) in SSPs.
- Optical CD-ROM, DVD-RAM, and DVD-ROM.
- Optical DVD-RAM backed by file.
- Tape devices.
- USB mass storage devices.

vSCSI is based on a client/server relationship model, as described in the following points.

- The VIOS owns the physical resources and the vSCSI server adapter, and acts as a server, or SCSI target device. The client LPARs have a SCSI initiator that is referred to as the vSCSI client adapter, and accesses the vSCSI targets as standard SCSI LUNs.
- Virtual disk resources can be configured and provisioned by using the HMC or the VIOS command-line interface (CLI).
- Physical disks that are owned by the VIOS can be exported and assigned to a client LPAR as a whole, added to an SSP, or partitioned into parts, such as logical volumes or files. Then, the logical volumes and files can be assigned to different LPARs. Therefore, by using vSCSI, you can share adapters and disk devices.
- LUs in logical volumes and file-backed virtual devices prevent the client partition from participating in LPM. To make a physical volume, logical volume, or file available to a client LPAR requires that it must be assigned to a vSCSI server adapter on the VIOS. The client LPAR accesses its assigned disks through a vSCSI client adapter. The vSCSI client adapter recognizes standard SCSI devices and LUNs through this virtual adapter.

For more information about vSCSI, see Planning for virtual SCSI, found at:

https://www.ibm.com/docs/en/power10/9080-HEX?topic=overview-virtual-scsi

Performance considerations

If sufficient CPU processing capacity is available, the performance of vSCSI must be comparable to dedicated I/O devices.

Virtual Ethernet, which has nonpersistent traffic, runs at a higher priority than the vSCSI on the VIOS. To make sure that high volumes of networking traffic do not starve vSCSI of CPU cycles, a threaded mode of operation is implemented for the VIOS by default since Version 1.2.

For more information about performance differences between physical and virtual I/O, see Planning for virtual SCSI, found at:

https://www.ibm.com/docs/en/power10/9080-HEX?topic=planning-virtual-scsi

Maximum number of slots

vSCSI itself does not have any maximums in terms of number of supported devices or adapters. The VIOS supports a maximum of 1024 virtual I/O slots per VIOS. A maximum of 256 virtual I/O slots can be assigned to a single client partition.

Every I/O slot needs some physical server resources to be created. Therefore, the resources that are assigned to the VIOS put a limit on the number of virtual adapters that can be configured.

For more information about limitation restrictions, see Limitations and restrictions of the Virtual I/O Server configuration, found at:

https://www.ibm.com/docs/en/power10/9080-HEX?topic=planning-limitations-restrictions-virtual-io-server-configuration

Naming conventions

A well-planned naming convention is key in managing the information. One strategy for reducing the amount of data that must be tracked is to make settings match on the virtual I/O client and server wherever possible.

The naming convention might include corresponding volume group, logical volume, and virtual target device (VTD) names. Integrating the virtual I/O client hostname into the VTD name can simplify tracking on the server.

Virtual device slot numbers

All vSCSI and Virtual Ethernet devices have slot numbers. In complex systems, there tends to be far more storage devices than network devices because each vSCSI device can communicate only with one server or client.

As shown in Figure 3-7 on page 103, the default value is 10 when you create an LPAR. The appropriate number for your environment depends on the number of virtual servers and adapters that are expected on each system. Each unused virtual adapter slot consumes a small amount of memory, so the allocation must be balanced. It is a best practice to set the maximum virtual adapters number to at least 100 on VIOS.

Important: When you plan for the number of virtual I/O slots on your LPAR, the maximum number of virtual adapter slots that is available on a partition is set by the partition's profile. To increase the maximum number of virtual adapters, you must change the profile, stop the partition (not just a restart), and start the partition.

To add virtual I/O clients without shutting down the LPAR or VIOS partition, leave plenty of room for expansion when the maximum number of slots are set.

The maximum number of virtual adapters must not be set higher than 1024 because that setting can cause performance problems.

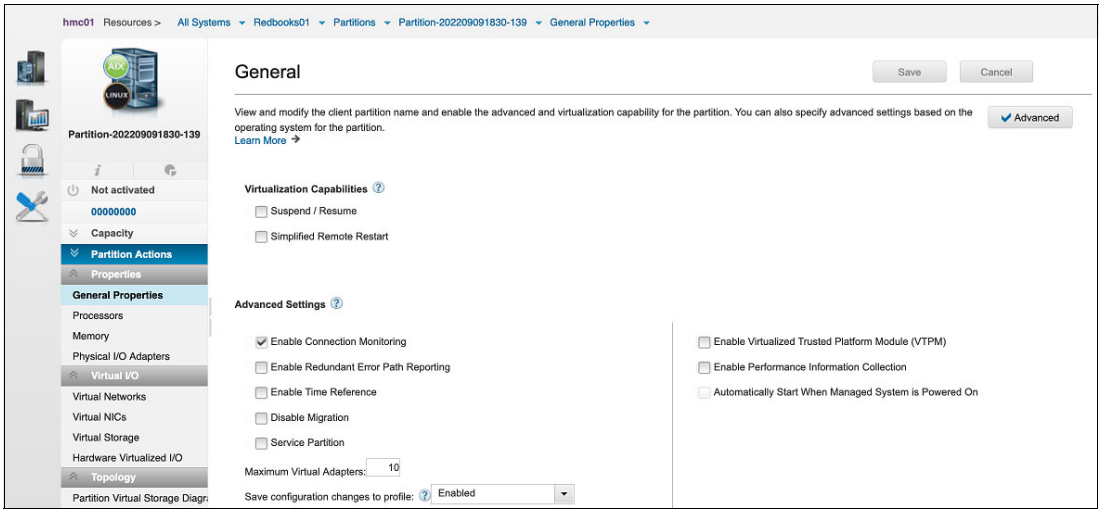


Figure 3-7 Setting the maximum limits in the partition's properties

For AIX virtual I/O client partitions, each adapter pair can handle up to 85 virtual devices with the default queue depth of three.

For IBM i clients, up to 16 virtual disk and 16 optical devices are supported.

For Linux clients, by default, up to 192 vSCSI targets are supported.

In situations where virtual devices per partition are expected to exceed these numbers, or where the queue depth on certain devices might be increased over the default, reserve extra adapter slots for the VIOS and the virtual I/O client partition.

When queue depths are tuned, the vSCSI adapters have a fixed queue depth. There are 512 command elements, of which two are used by the adapter, three are reserved for each vSCSI LUN for error recovery, and the rest are used for I/O requests. Thus, the default queue depth of 3 for vSCSI LUNs allows for up to 85 LUNs to use an adapter: (512 - 2) / (3 + 3) = 85 rounding down. If you need higher queue depths for the devices, the number of LUNs per adapter is reduced. For example, if you want to use a queue depth of 25, it allows 510/28 = 18 LUNs per adapter for an AIX client partition.

For Linux clients, the maximum number of LUNs per vSCSI adapter is decided by the **`max_id`** and **`max_channel`** parameters. The **`max_id`** is set to 3 by default, which can be increased to 7. The **`max_channel`** parameter is set to 64 by default, which is the maximum value. With the default values, the Linux client can have 3 * 64 = 192 vSCSI targets. If you overload an adapter, your performance is reduced.

Adding multiple adapters between a VIOS and a client must be considered when you are using mirroring on the virtual I/O client across multiple storage subsystems for availability.

For more information about capacity planning for latency, bandwidth, and sizing considerations, see Planning for virtual SCSI, found at:

https://www.ibm.com/docs/en/power10/9080-HEX?topic=planning-virtual-scsi

Virtual SCSI limitations for IBM i

The vSCSI limitations for IBM i are as follows:

- The IBM i 7.1 TR8 or later client LPARs can have up to 32 disk units (logical volumes, physical volumes, or files) and up to 16 optical units under a single virtual adapter.
- The maximum virtual disk size is 2 TB minus 512 bytes. If you are limited to one adapter and you have a storage requirement of 32 TB, for example, you might need to make your virtual disks the maximum size of 2 TB. However, in general, consider spreading the storage over multiple virtual disks with smaller capacities. This approach can help improve concurrency.
- Mirroring and multipath through up to eight VIOS partitions is the redundancy option for client LPARs. However, you also can use multipathing and RAID on the VIOS for redundancy.
- You must assign the tape device to its own VIOS adapter because tape devices often send large amounts of data that might affect the performance of any other device on the adapter.

For more information, see Multipathing and disk resiliency with vSCSI in a dual VIOS configuration, found at:

https://www.ibm.com/support/pages/multipathing-and-disk-resiliency-vscsi-dual-vios-configuration

3.5.2 Virtual Fibre Channel planning

N_Port ID Virtualization (NPIV) is an industry-standard technology that helps you to configure an NPIV-capable FC adapter with multiple, virtual worldwide port names (WWPNs). This technology is also called VFC. Similar to the virtual vSCSI function (vSCSI), VFC is a method to securely share a physical FC adapter among multiple VIOSs.

From an architectural perspective, the key difference between VFC and vSCSI is that the VIOS does not act as a SCSI emulator to its client partitions. Instead, it acts as a direct FC pass-through for the Fibre Channel Protocol (FCP) I/O traffic through the hypervisor. The client partitions are presented with full access to the physical SCSI target devices of a SAN disk or tape storage systems. The benefits of VFC are that the physical target device characteristics such as vendor or model information remains fully visible to the VIOS. Hence, you do not change the device drivers such as multi-pathing software, middleware such as copy services, or storage management applications that rely on the physical device characteristics.

For each VFC client adapter, two unique, virtual WWPNs, starting with the letter *c,* are generated by the HMC. After the activation of the client partition, the WWPNs log in to the SAN similar to other WWPNs from a physical port.

NPIV is described in 2.3.2, “Virtual Fibre Channel” on page 43.

Role of Virtual I/O Server

For VFC, the VIOS acts as an FC pass-through instead of a SCSI emulator, such as when vSCSI is used. A comparison between vSCSI and VFC is shown in Figure 3-8.

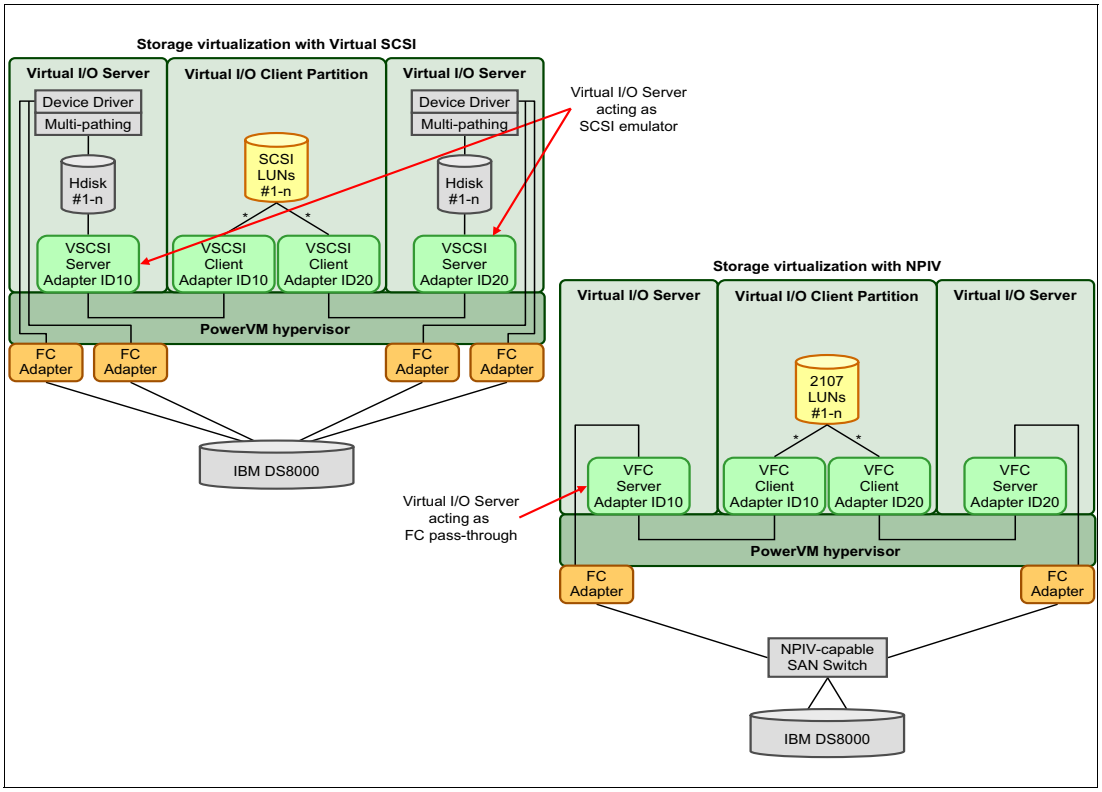


Figure 3-8 Comparing virtual SCSI and Virtual Fibre Channel

Two unique virtual WWPNs starting with the letter “c” are generated by the HMC for the VFC client adapter. After activation of the client partition, these WWPNs log in to the SAN like any other WWPNs from a physical port. Therefore, disk or tape storage target devices can be assigned to them as though they were physical FC ports.

Planning considerations for Virtual Fibre Channel

Consider the following information when you use the VFC:

- One VFC client adapter per physical port per client partition. This strategy helps to avoid a single point of failure (SPOF).
- For 16 GBps or slower FC adapters, a maximum of 64 active VFC client adapters per physical port. The virtual adapters per physical port can be reduced due to other VIOS resource constraints.
- For 32 GBps or faster FC adapters, a maximum of 255 VFC client adapters per physical port. The virtual adapters per physical port can be reduced because of other VIOS resource constraints.
- Maximum of 64 targets per VFC adapter.

- 32,000 unique WWPN pairs per system. Removing a VFC client adapter does not reclaim WWPNs. You can manually reclaim WWPNs by using the **mksyscfg** and **chhwres** commands or by using the **virtual_fc_adapters** attribute.
- To enable VFC on the managed system, create the required VFC adapters and connections by using HMC, as described in Chapter 4, “Implementing IBM PowerVM” on page 139.

The HMC generates WWPNs based on the range of names that is available for use with the prefix in the vital product data on the managed system. You can get the 6-digit prefix when you purchase the managed system. The 6-digit prefix includes 32,000 pairs of WWPNs. When you remove a VFC adapter from a client partition, the hypervisor deletes the WWPNs that are assigned to the VFC adapter on the client partition. The HMC does not reuse the deleted WWPNs to generate WWPNs for VFC adapters. If you require more WWPNs, you must obtain an activation code that includes another prefix that has another 32,000 pairs of WWPNs.

To avoid configuring the physical FC adapter to be a SPOF for the connection between the client partition and its physical storage on the SAN, do not connect two VFC adapters from the same client partition to the same physical FC adapter. Instead, connect each VFC adapter to a different physical FC adapter.

On a server that is managed by the HMC, you can dynamically add and remove VFC adapters to and from the VIOS and from each client partition. You can also view information about the virtual and physical FC adapters and the WWPNs by using VIOS commands.

Virtual Fibre Channel limitations for IBM i

The VFC limitations for IBM i are as follows:

- The IBM i client partition supports up to 128 target port connections per VFC adapter.
- The IBM i 7.3 and IBM i 7.4 client partitions support up to 127 SCSI devices per VFC adapter. The 127 SCSI devices can be any combination of disk units or tape libraries. With tape libraries, each control path is counted as a unique SCSI device in addition to a single SCSI device per tape drive.
- For IBM i client partitions, the LUNs of the physical storage that is connected with VFC require a storage-specific device driver and do not use the generic vSCSI device driver.
- The IBM i client partition supports up to eight multipath connections to a single FC disk unit. Each multipath connection can be made with a VFC adapter or with FC I/O adapter hardware that is assigned to the IBM i partition.
- IBM i supports mapping the same physical FC port to multiple VFC adapters in the same IBM i client. All LUNs (disk or tape) that are associated to that physical FC adapter must be unique so that no multi-path is created within the same physical port. To use LPM or remote restart capability, you can map only the physical port twice to the same IBM i LPAR. The VIOS must be at the Version 3.1.2.0 or later. The HMC must be at Version 9.2.950 or later. These versions are required for the LPM and to restart the LPAR with double-mapped ports support.
- With VIOS, you can install IBM i in a client LPAR on Power9 or Power10 processor-based systems. IBM i client LPARs have unique system and storage requirements and considerations.

For more information about VFC limitations for IBM i, see Limitations and restrictions for IBM i client logical partitions, found at:

https://www.ibm.com/docs/en/power10/9080-HEX?topic=planning-i-restrictions

3.5.3 Redundancy configurations for virtual Fibre Channel adapters

To implement highly reliable virtual I/O storage configurations, plan the following redundancy configurations to protect your virtual I/O production environment from physical adapter failures and from VIOS failures.

With NPIV, you can configure the managed system so that multiple LPARs can access independent physical storage through the same physical FC adapter. Each VFC adapter is identified by a unique WWPN, which means that you can connect each VFC adapter to independent physical storage on a SAN.

Host bus adapter redundancy

Similar to vSCSI redundancy, VFC redundancy can be achieved by using multipathing or mirroring at the client LPAR. The difference between redundancy with vSCSI adapters and the VFC technology that uses VFC client adapters is that the redundancy occurs at the client because only the virtual I/O client LPAR recognizes the disk. The VIOS is just an FC pass-through managing the data transfer through the hypervisor.

Host bus adapter (HBA) failover provides a basic level of redundancy for the client LPAR. Figure 3-9 shows the connectivity example.

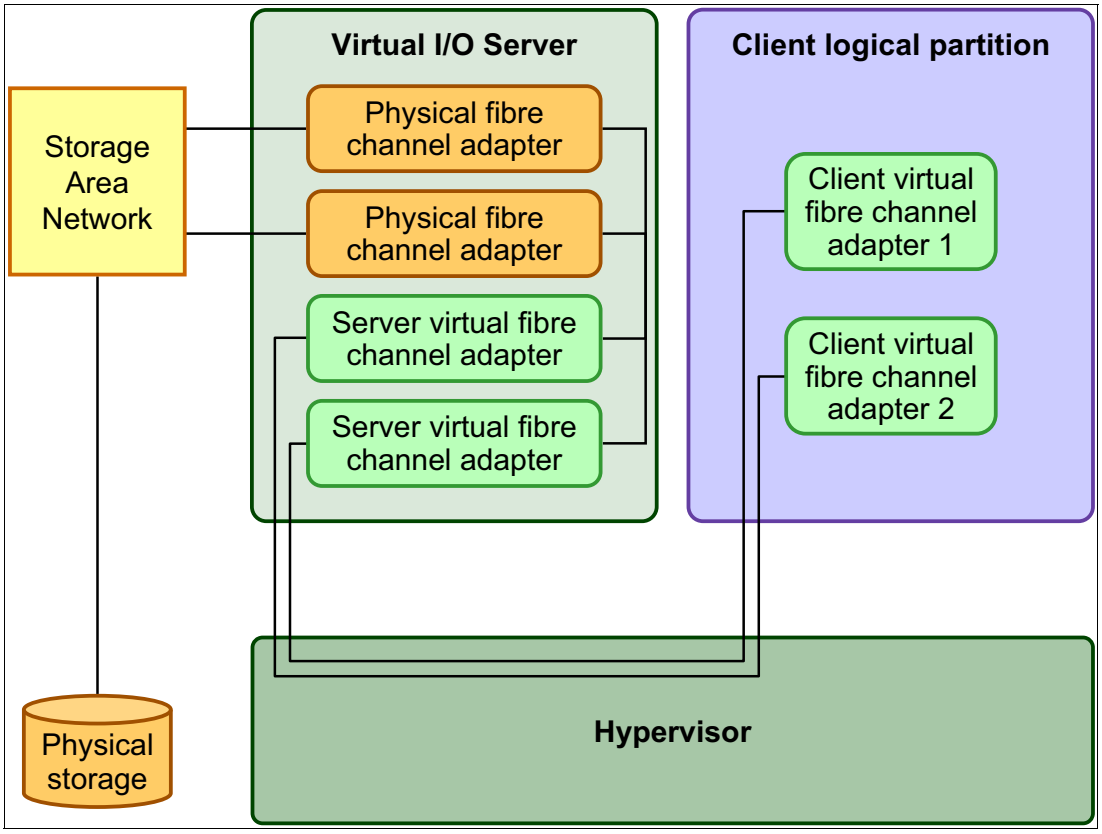


Figure 3-9 Host bus adapter connectivity

- The SAN connects physical storage to two physical FC adapters that are on the managed system.
- The physical FC adapters are assigned to the VIOS and support NPIV.

- The physical FC ports are each connected to a VFC adapter on the VIOS. The two VFC adapters on the VIOS are connected to ports on two different physical FC adapters to provide redundancy for the physical adapters.
- Each VFC adapter on the VIOS is connected to one VFC adapter on a client LPAR. Each VFC adapter on each client LPAR receives a pair of unique WWPNs. The client LPAR uses one WWPN to log in to the SAN at any specific time. The other WWPN is used when you move the client LPAR to another managed system.
- The VFC adapters always have a one-to-one relationship between the client LPARs and the VFC adapters on the VIOS LPAR. That is, each VFC adapter that is assigned to a client LPAR must connect to only one VFC adapter on the VIOS. Also, each VFC on the VIOS must connect to only one VFC adapter on a client LPAR.

Note: As a best practice, configure VFC adapters from multiple LPARs to the same HBA, or configure VFC adapters from the same LPAR to different HBAs.

Host bus adapter and Virtual I/O Server redundancy

An HBA and VIOS redundancy configuration provide a more advanced level of redundancy for the virtual I/O client partition. Figure 3-10 on page 109 shows the following connections:

- The SAN connects physical storage to two physical FC adapters that are on the managed system.
- Two VIOS LPARs provide redundancy at the VIOS level.
- The physical FC adapters are assigned to their respective VIOS and support NPIV.
- The physical FC ports are each connected to a VFC adapter on the VIOS.
- The two virtual FC adapters on the VIOS are connected to ports on two different physical FC adapters to provide redundancy for the physical adapters. A single adapter might have multiple ports.
- Each VFC adapter on the VIOS is connected to one VFC adapter on a client LPAR. Each VFC adapter on each client LPAR receives a pair of unique WWPNs. The client LPAR uses one WWPN to log in to the SAN at any specific time. The other WWPN is used when you move the client LPAR to another managed system.

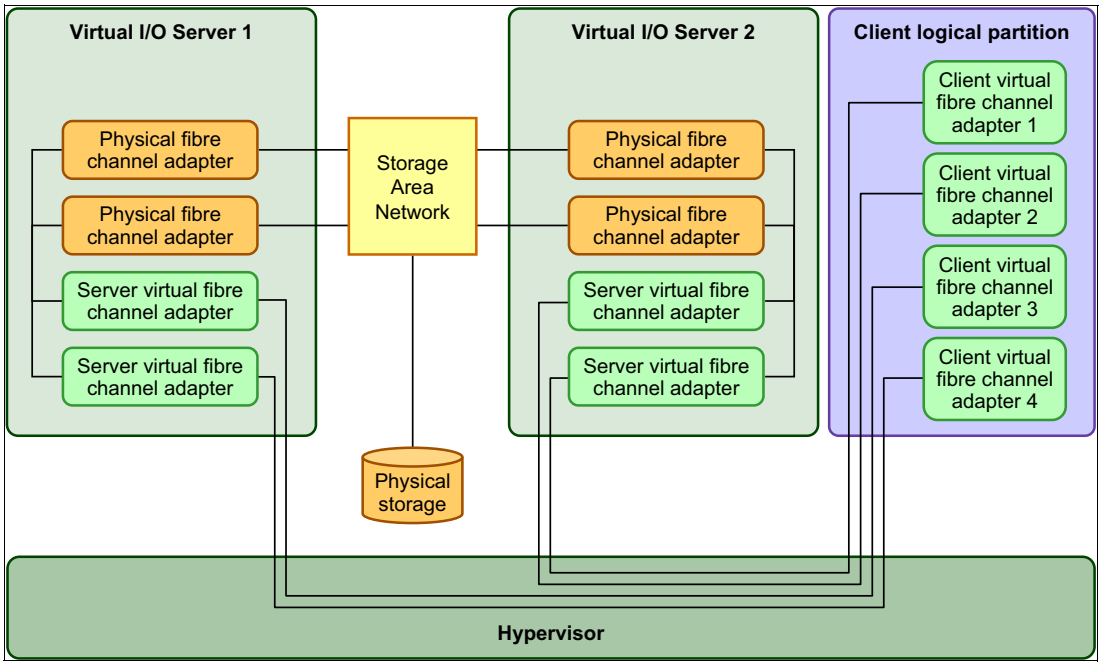


Figure 3-10 A host bus adapter and Virtual I/O Server redundancy

The client can write to the physical storage through VFC adapter 1 or 2 on the client LPAR through VIOS 2. The client also can write to physical storage through VFC adapter 3 or 4 on the client LPAR through VIOS 1. If a physical FC adapter fails on VIOS 1, the client uses the other physical adapter that is connected to VIOS 1 or uses the paths that are connected through VIOS 2. If VIOS 1 fails, then the client uses the path through VIOS 2. This example does not show redundancy in the physical storage, but assumes it is built into the SAN.

Other considerations for virtual Fibre Channel

These examples can become more complex as you add physical storage redundancy and multiple clients, but the concepts remain the same. Consider the following points:

- To avoid configuring the physical FC adapter to be a SPOF for the connection between the client LPAR and its physical storage on the SAN, do not connect two VFC adapters from the same client LPAR to the same physical FC adapter. Instead, connect each VFC adapter to a different physical FC adapter.
- Consider load-balancing when a VFC adapter on the VIOS is mapped to a physical port on the physical FC adapter.
- Consider what level of redundancy exists in the SAN to determine whether to configure multiple physical storage units.
- Consider the usage of two VIOS LPARs. Because the VIOS is central to communication between LPARs and the external network, it is important to provide a level of redundancy for the VIOS. Multiple VIOS LPARs require more resources too, so you must plan for them too.

Using their unique WWPNs and the VFC connections to the physical FC adapter, the client operating system that runs in the virtual I/O client partitions discovers, instantiates, and manages the physical storage that is on the SAN as though it were natively connected to the SAN storage device. The VIOS provides the virtual I/O client partitions with a connection to the physical FC adapters on the managed system.

A one-to-one relationship always exists between the VFC client adapter and the VFC server adapter.

The SAN uses zones to provide access to the targets based on WWPNs. VFC client adapters are created by using HMC with unique set of WWPNs. VFC adapters can be zoned for SAN access, just like physical FC adapters.

Redundancy configurations help to increase the serviceability of your VIOS environment. With VFC, you can configure the managed system so that multiple virtual I/O client partitions can independently access physical storage through the same physical FC adapter. Each VFC client adapter is identified by a unique WWPN, which means that you can connect each virtual I/O partition to independent physical storage on a SAN.

Mixtures: Though any mixture of VIOS native SCSI, vSCSI, and VFC I/O traffic is supported on the same physical FC adapter port, consider the implications that this mixed configuration might have for manageability and serviceability.

IBM i Virtual Fibre Channel recommendations

For more information about recommendations that can be followed to ensure that VFC and NPIV environments perform as well as possible when they are connected to supported external storage systems, see IBM i Virtual Fibre Channel Performance Best Practices, found at:

https://www.ibm.com/support/pages/ibm-i-virtual-fibre-channel-performance-best-practices

For more information about Virtual Fibre Channel planning, see How to prepare for SAN changes in a Virtualized Fibre Channel NPIV environment, found at:

https://www.ibm.com/support/pages/node/6610641

For more information about recommended device attributes for redundancy, see Configuring a VIOS for client storage, found at:

https://www.ibm.com/docs/en/power10?topic=partition-configuring-vios-client-storage

3.5.4 Virtual SCSI and Virtual Fibre Channel comparison

vSCSI and VFC both offer significant benefits by enabling shared utilization of physical I/O resources. The following sections compare both capabilities and provide guidance for selecting the most suitable option.

Overview

Table 3-4 shows a high-level comparison of vSCSI and VFC.

Table 3-4 Virtual SCSI and Virtual Fibre Channel comparison

Feature	Virtual SCSI	VFC
Server-based storage virtualization	Yes	No
Adapter-level sharing	Yes	Yes
Device-level sharing	Yes	No
LPM-capable	Yes	Yes

Feature	Virtual SCSI	VFC
SSP-capable	Yes	No
SCSI-3 compliant (persistent reserve)	No[a]	Yes
Generic device interface	Yes	No
Tape library and LAN-free backup support	No	Yes
Virtual tape and virtual optical support	Yes	No
Support for IBM PowerHA System Mirror for i[b]	No	Yes

a. Unless using SSPs.
b. Applies only to IBM i partitions.

Components and features

The following section describes the various components and features.

Device types

vSCSI provides virtualized access to disk devices, optical devices, and tape devices. With VFC, SAN disk devices and tape libraries can be attached. The access to tape libraries enables the usage of LAN-free backup, which is not possible with vSCSI.

Adapter and device sharing

vSCSI allows sharing of physical storage adapters. It also allows sharing of storage devices by creating storage pools that can be partitioned to provide logical volume or file-backed devices.

VFC technology allows sharing of physical FC adapters only.

Hardware requirements

VFC implementation requires VFC-capable FC adapters on the VIOS and VFC-capable SAN switches.

vSCSI supports a broad range of physical adapters.

Storage virtualization

vSCSI server provides servers-based storage virtualization. Storage resources can be aggregated and pooled on the VIOS.

When VFC is used, the VIOS is only passing-through I/O to the client partition. Storage virtualization is done on the storage infrastructure in the SAN.

Storage assignment

With vSCSI, the storage is assigned (zoned) to the VIOSs. From a storage administration perspective, no end-to-end view to see which storage is allocated to which client partition is available. When new disks are added to an existing client partition, they must be mapped on the VIOS. When LPM is used, storage must be assigned to the VIOSs on the target server.

With VFC, the storage is assigned to the client partitions, as in an environment where physical adapters are used. No intervention is required on the VIOS when new disks are added to an existing partition. When LPM is used, storage moves to the target server without requiring a reassignment because the VFCs have their own WWPNs that move with the client partitions to the target server.

Support of PowerVM capabilities

Both vSCSI and VFC support most PowerVM capabilities, such as LPM.

VFC does not support virtualization capabilities that are based on the SSP, such as thin-provisioning.

Client partition considerations

vSCSI uses a generic device interface, which means regardless of the backing device that is used, the devices appear in the same way in the client partition. When vSCSI is used, no additional device drivers must be installed in the client partition. vSCSI does not support load-balancing across virtual adapters in a client partition.

With VFC, a tape device driver must be installed in the client partition for the disk devices or tape devices. Native AIX MPIO allows load-balancing across virtual adapters. Upgrading these drivers requires special attention when you use SAN devices as boot disks for the operating system.

Worldwide port names

With the redundant configurations that use two VIOSs and two physical FC adapters that are explained in 3.5.3, “Redundancy configurations for virtual Fibre Channel adapters” on page 107, up to eight WWPNs are used. Some SAN storage devices have a limit on the number of WWPNs that they can manage. Therefore, before VFC is deployed, verify that the SAN infrastructure can support the planned number of WWPNs. vSCSI uses only WWPNs of the physical adapters on the VIOS.

Hybrid configurations

vSCSI and VFC can be deployed in hybrid configurations. The next two examples show how both capabilities can be combined in real-world scenarios:

1. In an environment that is constrained by the number of WWPNs, vSCSI can be used to provide access to disk devices.
2. For partitions that require LAN-free backup, access to tape libraries can be provided by using VFC.

To simplify the upgrade of device drivers, VFC can be used to provide access to application data, and vSCSI can be used for access to the operating system boot disks.

3.5.5 Availability planning for virtual storage

This section provides planning details that are required to set up redundancy for virtual storage.

Virtual storage redundancy

VFC or vSCSI redundancy can be achieved by using MPIO and LVM mirroring at the client partition and VIOS level.

Figure 3-10 on page 109 depicts a redundant VFC configuration. Review the description under that figure to understand how to implement highly reliable virtual I/O storage configurations that are based on VFC technology.

Figure 3-11 depicts a vSCSI redundancy advanced setup by using both MPIO and LVM mirroring in the client partition concurrently, with two VIOSs host disks for a client partition. The client is using MPIO to access a SAN disk and LVM mirroring to access two SCSI disks. From the client perspective, the following situations can be handled without causing downtime for the client:

- Either path to the SAN disk can fail, but the client still can access the data on the SAN disk through the other path. No actions must be taken to reintegrate the failed path to the SAN disk after repair if MPIO is configured.
- The failure of a SCSI disk causes stale partitions on AIX for the volume group with the assigned virtual disks, a suspended disk unit on IBM i, or a disk marked as failed on Linux. The client partition still can access the data on the second copy of the mirrored disk. After the failed disk is available again, the stale partitions must be synchronized on the AIX client by using the **varyonvg** command. The IBM i client automatically resumes mirrored protection, while on the Linux client, the command **mdadm** and a rescan of the devices are required.
- Either VIOS can be restarted for maintenance. This action results in a temporary simultaneous failure of one path to the SAN disk and stale partitions for the volume group on the SCSI disks, as described before.

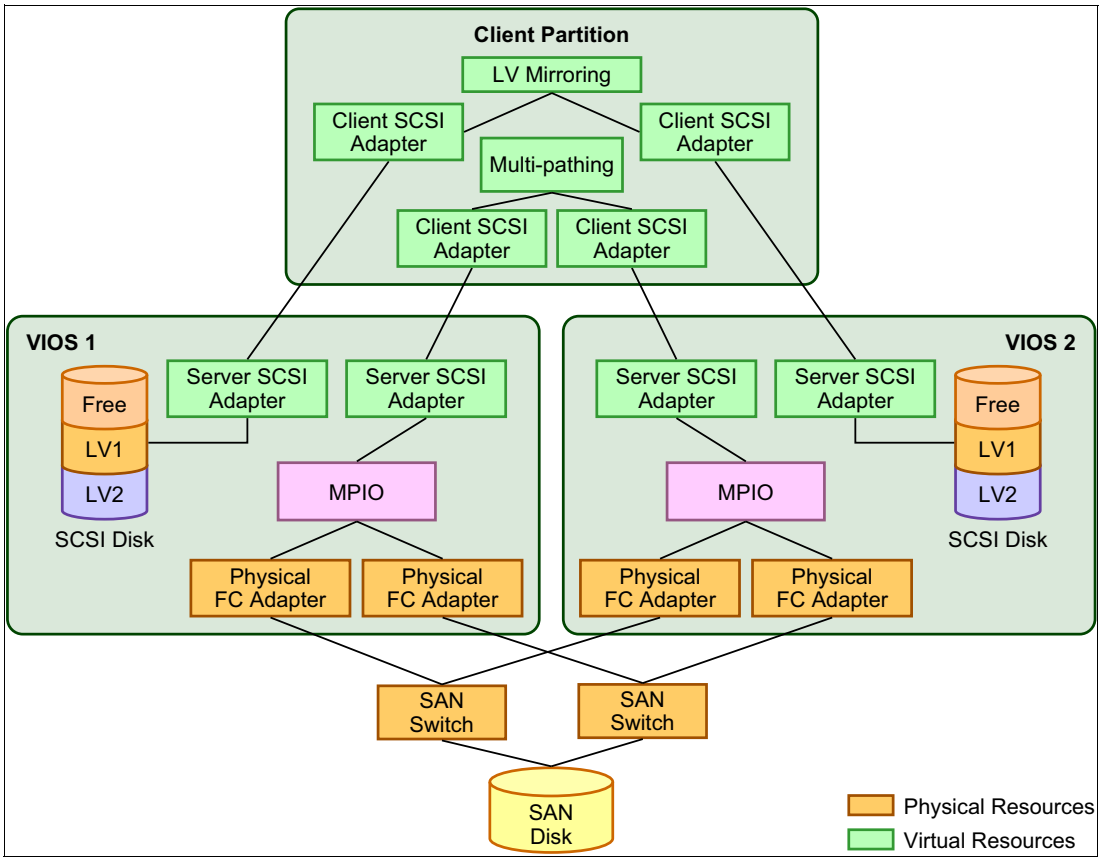


Figure 3-11 Virtual SCSI redundancy by using multipathing and mirroring

Considerations for redundancy

Consider the following points:

- If mirroring and multipathing are both configurable in your setup, multipathing is the preferred method for adding disk connection redundancy to the client. Mirroring causes stale partitions on AIX or Linux, and suspended disk units on IBM i, which require synchronization, but multipathing does not. Depending on the RAID level that is used on the SAN disks, the disk space requirements for mirroring can be higher. Mirroring across two storage systems even allows enhancement of the redundancy that is provided in a single storage system by RAID technology.
- Two FC adapters in each VIOS allow for adapter redundancy.

The following sections describe the usage of mirroring for each different AIX, IBM i, and Linux client partition across two VIOSs.

AIX LVM mirroring in the client partition

To provide storage redundancy in the AIX client partition, AIX LVM mirroring can be used for VFC devices or vSCSI devices.

When vSCSI and AIX client partition mirroring is used between two storage subsystems, in certain situations, errors on hdisks that are on a single storage subsystem can cause all hdisks that are connected to a vSCSI adapter to become inaccessible.

To avoid losing access to mirrored data, a best practice is to provide the disks of each mirror copy through a different vSCSI adapter, as shown in Figure 3-12 on page 115.

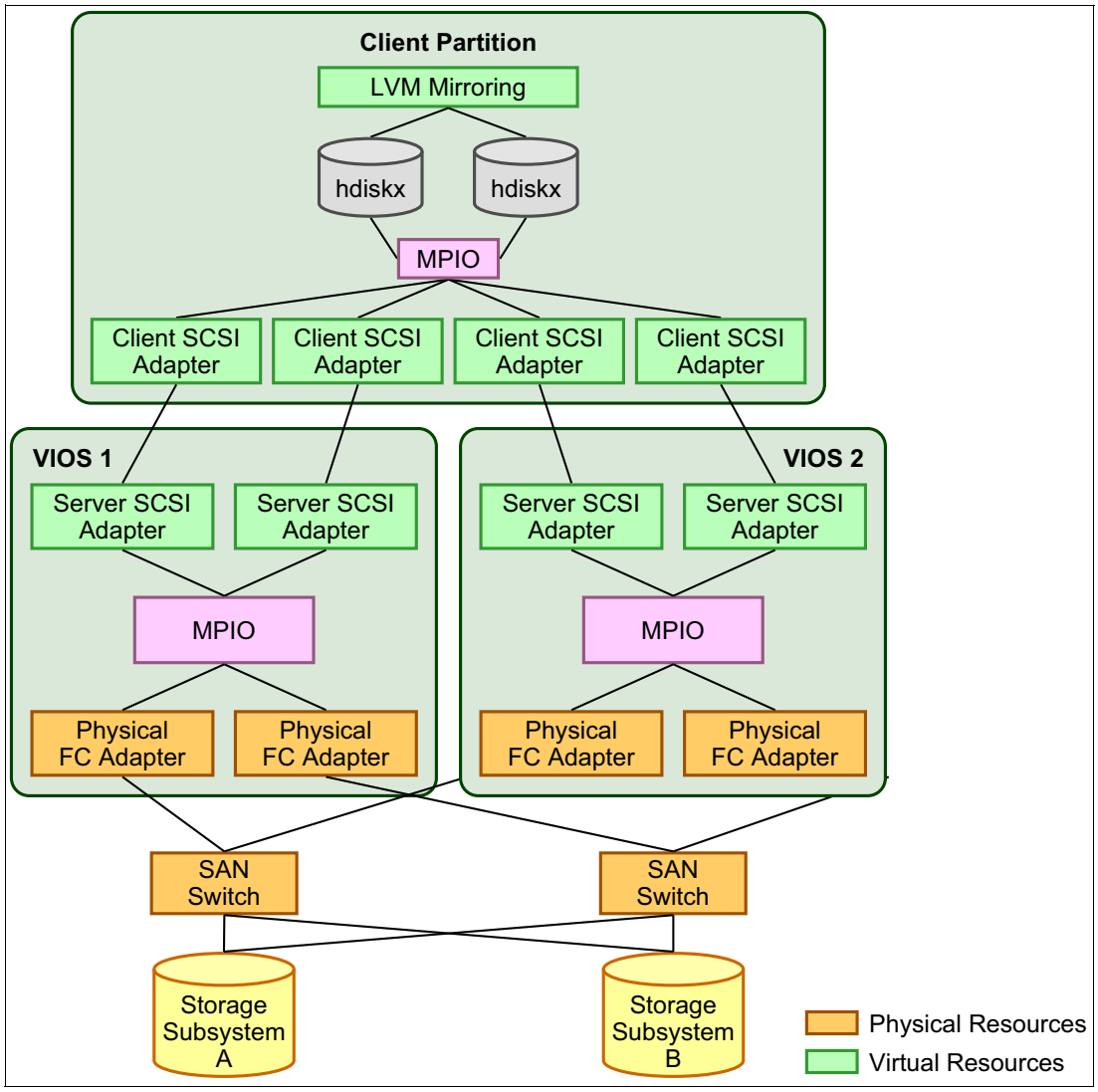


Figure 3-12 LVM mirroring with two storage subsystems

Volume group mirroring in the AIX client is also a best practice when a logical volume in VIOS is used as a vSCSI device on the client. In this case, the vSCSI devices are associated with different SCSI disks, each one controlled by one of the two VIOS, as shown in Figure 3-12. Mirroring logical volumes in VIOS is not necessary when the data is mirrored in the AIX client.

IBM i mirroring in the client partition

IBM i mirroring in the client partition to enable storage redundancy, ideally across two VIOSs and two separate storage systems, is supported for vSCSI or IBM DS8000® VFC LUNs that are attached by VFC.

vSCSI LUNs are presented by the VIOS as unprotected LUNs of type-model 6B22-050 to the IBM i client so they are eligible for IBM i mirroring. For DS8000 series VFC LUNs, as with DS8000 series native attachment, the LUNs must be created as unprotected models (IBM OS/400® model A8x) on the DS8000 series to be eligible for IBM i mirroring.

Important: Currently, all vSCSI or FC adapters report on IBM i under the same bus number 255, which allows for IOP-level mirrored protection only. To implement the concept of bus-level mirrored protection for virtual LUNs with larger configurations with more than one virtual IOP per mirror side and not compromise redundancy, consider iteratively adding LUNs from one IOP pair at a time to the auxiliary storage pool by selecting the LUNs from one virtual IOP from each mirror side.

Linux mirroring in the client partition

Mirroring on Linux partitions is implemented with a Linux software RAID function that is provided by an **md** (Multiple Devices) device driver. The **md** driver combines devices in one array for performance improvements and redundancy.

An **md** device with RAID 1 indicates a mirrored device with redundancy. RAID devices on Linux are represented as md0, md1, and so on.

Linux software RAID devices are managed and listed with the **mdadm** command. You also can list RAID devices with the **cat /proc/mdstat** command.

All devices in a RAID1 array must have the same size; otherwise, the smallest device space is used, and any extra space on other devices is wasted.

3.5.6 Shared storage pools planning

This section describes the necessary planning details for implementing SSPs in a PowerVM environment.

SSPs are described in 2.3.3, “Shared storage pools” on page 44.

The following sections list the prerequisites for creating SSPs.

Prerequisites

Ensure that the following prerequisites are met:

- VIOS
- HMC
- Minimum 20 GB of available storage space for a storage pool
- Storage requirements of your storage vendor

Configuring the Virtual I/O Server logical partitions

Configure the VIOS LPARs as follows:

- There must be at least one CPU and one physical CPU of entitlement.
- The LPARs must be configured as VIOS LPARs.
- The LPARs must consist of at least 4 GB of memory.
- The LPARs must consist of at least one physical FC adapter.
- The rootvg device for a VIOS LPAR cannot be included in storage pool provisioning.
- The VIOS LPARs in the cluster require access to all the SAN-based physical volumes in the SSP of the cluster.

Scalability limits

The scalability limits of SSP Cluster on VIOS 3.1.3.0 are shown in the following fields:

- Max number of Nodes in cluster: 16
- Max Number of Physical Disks in Pool: 1024
- Max Number of Virtual Disks: 8192
- Max Number of Client LPARs per VIOS: 250 (requires that each VIOS has at least 4 CPUs and 8 GB memory)
- Max Capacity of Physical Disks in Pool: 16 TB
- Min/Max Storage Capacity of Storage Pool: 512 TB
- Max Capacity of a Virtual Disk (LU) in Pool: 4 TB

Configuring client logical partitions

Configure the client partitions with the following characteristics:

- The client LPARs must be configured as AIX or Linux client systems.
- They must have at least 1 GB of minimum memory.
- The associated rootvg device must be installed with the appropriate AIX or Linux system software.
- Each client LPAR must be configured with enough vSCSI adapter connections to map to the virtual server SCSI adapter connections of the required VIOS LPARs.

Network addressing considerations

Uninterrupted network connectivity is required for SSP operations. The network interface that is used for the SSP configuration must be on a highly reliable network, which is not congested.

Ensure that both the forward and reverse lookup for the hostname that is used by the VIOS LPAR for clustering resolves to the same IP address.

Notes:

- The SSP cluster can be created on an IPv6 configuration. Therefore, VIOS LPARs in a cluster can have hostnames that resolve to an IPv6 address. To set up an SSP cluster on an IPv6 network, IPv6 stateless autoconfiguration is suggested. You can have VIOS LPARs that are configured with either an IPv6 static configuration or an IPv6 stateless autoconfiguration. A VIOS that has both IPv6 static configuration and IPv6 stateless autoconfiguration is not supported.
- The hostname of each VIOS LPAR that belongs to the same cluster must resolve to the same IP address family, which is either IPv4 or IPv6.
- To change the hostname of a VIOS LPAR in the cluster, you must remove the partition from the cluster and change the hostname. Later, you can add the partition back to the cluster again with new hostname.
- Commands on VIOS (**`mktcpip`**, **`rmtcpip`**, **`chtcpip`**, **`hostmap`**, **`chdev`**, and **`rmdev`**) are enhanced to configure more than one network interface without disturbing its existing network configuration. In the SSP environment, this feature helps the user to configure multiple network interfaces without causing any harm to the existing SSP setup. In the presence of multiple network interfaces, the primary interface might not be the interface that is used for cluster communication. In such an SSP environment, the user is not restricted from altering the network configuration of other interfaces.

Storage provisioning to Virtual I/O Server partitions

When a cluster is created, you must specify one physical volume for the repository disk and at least one physical volume for the storage pool. The storage pool physical volumes are used to provide storage to the data that is generated by the client partitions. The repository disk is used to perform cluster communication and store the cluster configuration. The maximum client storage capacity matches the total storage capacity of all storage pool physical volumes. The repository disk must have at least 10 GB of available storage space. The physical volumes in the storage pool must have at least 20 GB of available storage space in total.

Use any method that is available for the SAN vendor to create each physical volume with at least 20 GB of available storage space. Map the physical volume to the LPAR's FC adapter for each VIOS in the cluster. The physical volumes must be mapped only to the VIOS LPARs that are connected to the SSP.

Note: Each of the VIOS LPARs assigns hdisk names to all physical volumes that are available through the FC ports, such as hdisk0 and hdisk1. The VIOS LPAR might select different hdisk numbers for the same volumes to the other VIOS LPAR in the same cluster. For example, the viosA1 VIOS LPAR can have hdisk9 assigned to a specific SAN disk, and the viosA2 VIOS LPAR can have the hdisk3 name assigned to that same disk. For some tasks, the unique device ID (UDID) can be used to distinguish the volumes. Use the **chkdev** command to obtain the UDID for each disk. It is also possible to rename the devices by using the **rendev** command.

Set the FC adapters parameters as follows:

```
chdev -dev fscsi0 -attr dyntrk=yes -perm
chdev -dev fscsi0 -attr fc_err_recov=fast_fail -perm
```

You do not need to set **no_reserve** on the repository disk or set it on any of the SSP disks. The Cluster Aware AIX (CAA) layer on the VIOS does this task.

3.6 Network virtualization planning

The following sections describe available network virtualization options and provide planning guidance for them.

3.6.1 Virtual Ethernet planning

Virtual Ethernet technology facilitates IP-based communication between LPARs on the same system by using virtual local area network (VLAN)-capable software switch systems. Using SEA technology, LPARs can communicate with other systems outside the hardware unit without being assigned physical Ethernet slots.

SEA is described in 2.4.2, "Shared Ethernet Adapter" on page 47.

You can create VEAs by using the HMC. You can add, remove, or modify the existing set of VLANs for a VEA that is assigned to an active partition by using the HMC.

Consider using virtual Ethernet on the VIOS in the following situations:

- When the capacity or the bandwidth requirement of the individual LPAR is inconsistent with or is less than the total bandwidth of a physical Ethernet adapter. LPARs that use the full bandwidth or capacity of a physical Ethernet adapter must either use SR-IOV technology or use dedicated Ethernet adapters.
- When you need an Ethernet connection, but no free slot is available where you can install a dedicated adapter.
- When advanced PowerVM virtualization technologies like LPM or partition Simplified Remote Restart (SRR) are used, you may *not* assign physical I/O devices to the client partitions. In this case, use virtual Ethernet with a SEA on the VIOS.

LPM is described in 2.6, “Partition mobility” on page 54, and SRR is described in 2.7, “Simplified Remote Restart” on page 59.

3.6.2 Virtual LAN planning

In many situations, the physical network topology must account for the physical constraints of the environment, such as rooms, walls, floors, and buildings.

However, VLANs can be independent of the physical topology. Figure 3-13 shows two VLANs (VLAN 1 and 2) that are defined on three switches (Switch A, B, and C). Seven hosts (A-1, A-2, B-1, B-2, B-3, C-1, and C-2) are connected to the three switches.

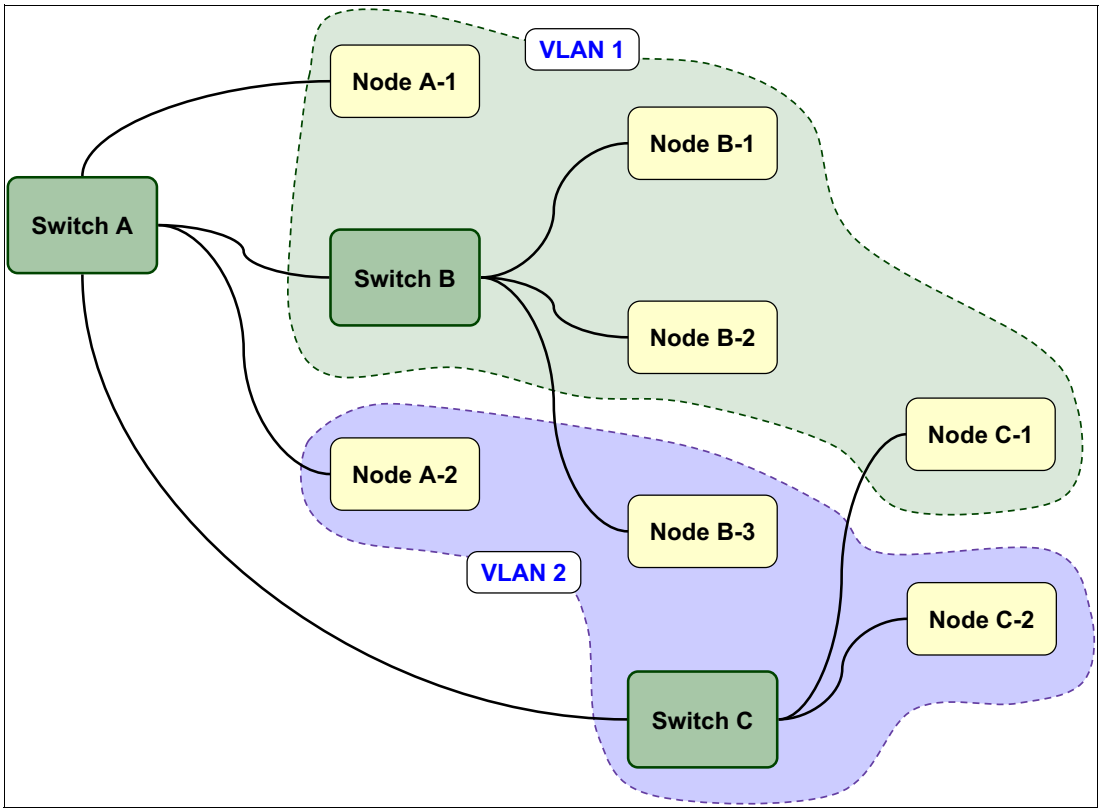


Figure 3-13 Multiple VLANs example

The physical network topology of the LAN forms a tree, which is typical for a nonredundant LAN:

- Switch A:
  - Node A-1
  - Node A-2
  - Switch B:
    - Node B-1
    - Node B-2
    - Node B-3
  - Switch C:
    - Node C-1
    - Node C-2

Although nodes C-1 and C-2 are physically connected to the same switch C, traffic between two nodes is blocked:

- VLAN 1:
  - Node A-1
  - Node B-1
  - Node B-2
  - Node C-1
- VLAN 2:
  - Node A-2
  - Node B-3
  - Node C-2

To enable communication between VLAN 1 and 2, L3 routing or inter-VLAN bridging must be established between the VLANs. The bridging is typically provided by an L3 device, for example, a router or firewall that is plugged into switch A.

Consider the uplinks between the switches. They carry traffic for both VLANs 1 and 2. Thus, there must be only one physical uplink from B to A, not one per VLAN. The switches are not confused and do not mix up the different VLANs' traffic because packets that travel through the trunk ports over the uplink are tagged.

VLANs also have the potential to improve network performance. By splitting up a network into different VLANs, you also split up broadcast domains. Thus, when a node sends a broadcast, only the nodes on the same VLAN are interrupted by receiving the broadcast. The reason is that normally broadcasts are not forwarded by routers. Consider this fact if you implement VLANs and want to use protocols that rely on broadcasting, such as Boot Protocol (BOOTP) or Dynamic Host Configuration Protocol (DHCP) for IP autoconfiguration.

It also is a best practice to use VLANs if gigabit Ethernet jumbo frames are implemented in an environment, where not all nodes or switches can use or are compatible with jumbo frames. Jumbo frames allow for a maximum transmission unit (MTU) size of 9000 instead of Ethernet’s default of 1500. This feature can improve throughput and reduce processor load on the receiving node in a heavy loaded scenario, such as backing up files over the network.

VLANs can provide extra security by allowing an administrator to block packets from one domain to another domain on the same switch. This approach provides more control over what LAN traffic is visible to specific Ethernet ports on the switch. Packet filters and firewalls can be placed between VLANs, and Network Address Translation (NAT) can be implemented between VLANs. VLANs can make the system less vulnerable to attacks.

3.6.3 Virtual switches planning

The PHYP switch is consistent with IEEE 802.1Q. It works on OSI-Layer 2 and supports up to 4094 networks (4094 VLAN IDs).

When a message arrives at a logical LAN switch port from a logical LAN adapter, the hypervisor caches the message's source MAC address to use as a filter for future messages to the adapter. Then, the hypervisor processes the message depending on whether the port is configured for IEEE VLAN headers. If the port is configured for VLAN headers, the VLAN header is checked against the port's allowable VLAN list. If the message-specified VLAN is not in the port's configuration, the message is dropped. After the message passes the VLAN header check, it passes to the destination MAC address for processing.

If the port is not configured for VLAN headers, the hypervisor inserts a 2-byte VLAN header (based on the VLAN number that is configured in the port) into the message. Next, the destination MAC address is processed by searching the table of cached MAC addresses.

If a match for the MAC address is not found and if no trunk adapter is defined for the specified VLAN number, the message is dropped. Otherwise, if a match for the MAC address is not found and if a trunk adapter is defined for the specified VLAN number, the message is passed on to the trunk adapter. If a MAC address match is found, then the associated switch port's allowable VLAN number table is scanned. It looks for a match with the VLAN number that is in the message's VLAN header. If a match is not found, the message is dropped.

Next, the VLAN header configuration of the destination switch port is checked. If the port is configured for VLAN headers, the message is delivered to the destination logical LAN adapters, including any inserted VLAN header. If the port is configured for no VLAN headers, the VLAN header is removed before it is delivered to the destination logical LAN adapter.

Figure 3-14 on page 122 shows a graphical representation of the behavior of the virtual Ethernet when processing packets.

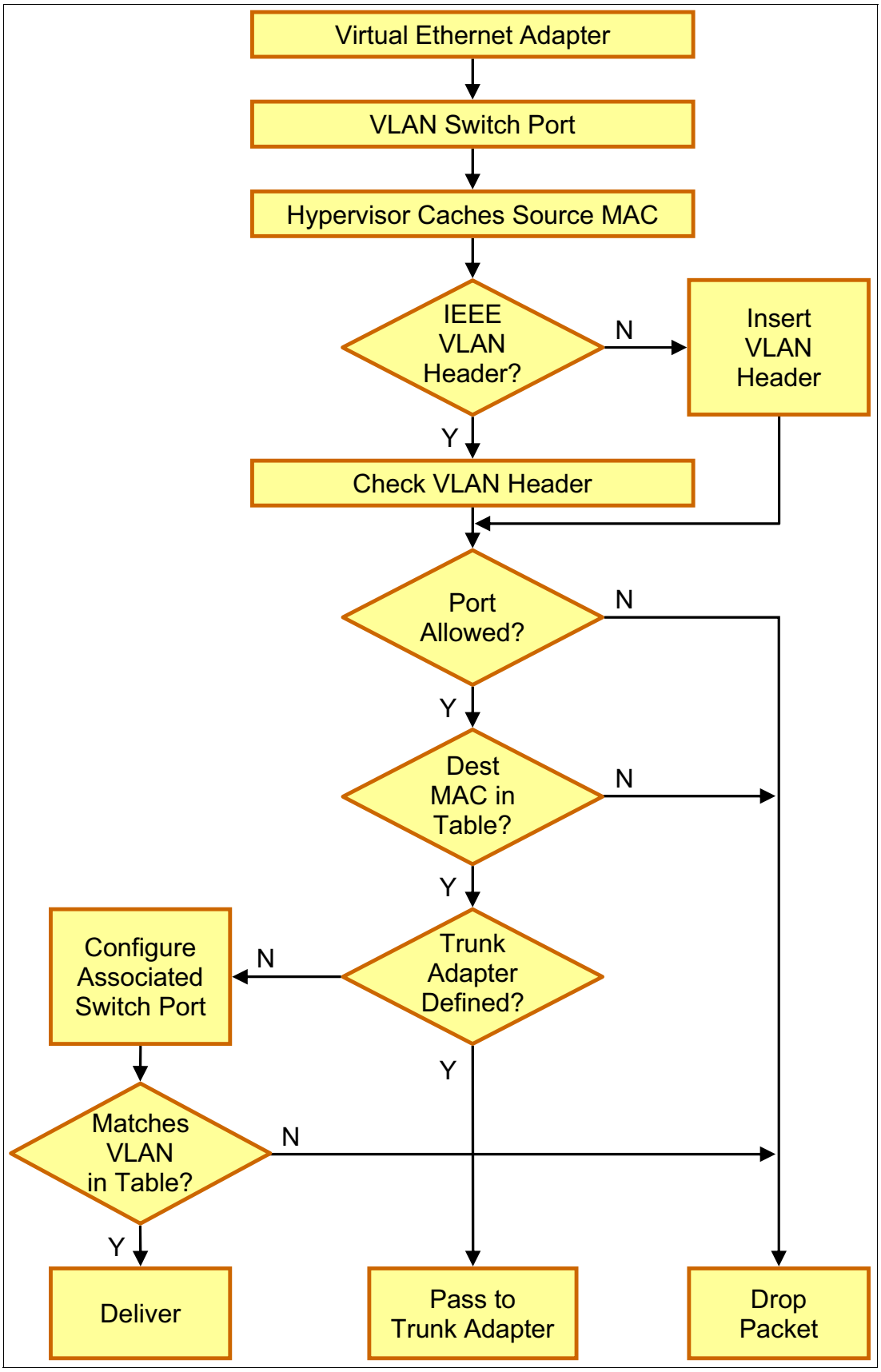


Figure 3-14 Flow chart of virtual Ethernet

Multiple virtual switches

Power servers support multiple virtual switches. By default, a single virtual switch that is named “Ethernet0” is configured. This name can be changed dynamically, and more virtual switches can be created with a name of your choice.

Extra virtual switches can be used to provide an extra layer of security or increase the flexibility of a virtual Ethernet configuration.

For example, to isolate traffic in a DMZ from an internal network without relying entirely on VLAN separation, two virtual switches can be used. The virtual adapters of the systems that participate in the DMZ network are configured to use one virtual switch, and systems that participate in the internal network are configured to use another virtual switch.

Consider the following points when multiple virtual switches are used:

- A VEA can be associated only with a single virtual switch.
- Each virtual switch supports the full range of VLAN IDs (1 - 4094).
- The same VLAN ID can exist in all virtual switches independently of each other.
- Virtual switches can be created and removed dynamically. However, a virtual switch cannot be removed if an active VEA is using it.
- Virtual switch names can be modified dynamically without interruption to connected VEAs.
- With LPM, virtual switch names must match between the source and target systems. The validation phase fails if names do not match.
- All virtual adapters in a SEA must be members of the same virtual switch.

Important: When a SEA is used, the name of the virtual switch is recorded in the configuration of the SEA on the VIOS at creation time. If the virtual switch name is modified, the name change is not reflected in this configuration until the VIOS is restarted, or the SEA device is reconfigured. The **`rmdev -l`** command followed by **`cfgmgr`** is sufficient to update the configuration. If the configuration is not updated, it can cause a Live Partition Migration validation process to fail because the VIOS still refers to the old name.

3.6.4 Shared Ethernet Adapter planning

SEA is described in 2.4.2, “Shared Ethernet Adapter” on page 47.

A SEA can be used to bridge a physical Ethernet network to a virtual Ethernet network. It also provides the ability for several client partitions to share one physical adapter. Using a SEA, you can connect internal and external VLANs by using a physical adapter. The SEA that is hosted in the VIOS acts as a layer-2 bridge between the internal and external network.

A SEA is a layer-2 network bridge to securely transport network traffic between virtual Ethernet networks and physical network adapters. The SEA service runs in the VIOS. It cannot be run in a general-purpose AIX or Linux partition.

Tip: A Linux partition also can provide a bridging function with the **`brctl`** command.

The SEA allows partitions to communicate outside the system without having to dedicate a physical I/O slot and a physical network adapter to a client partition. The SEA has the following characteristics:

- Virtual Ethernet MAC addresses of VEAs are visible to outside systems (by using the **arp -a** command).
- Unicast, broadcast, and multicast is supported. Therefore, protocols that rely on broadcast or multicast, such as Address Resolution Protocol (ARP), DHCP, BOOTP, and Neighbor Discovery Protocol (NDP) can work across an SEA.

To bridge network traffic between the virtual Ethernet and external networks, the VIOS must be configured with at least one physical Ethernet adapter. One SEA can be shared by multiple VEAs, and each one can support multiple VLANs. A SEA can include up to 16 VEAs on the VIOS that share the physical access.

Tip: An IP address does not need to be configured on a SEA to perform the Ethernet bridging function. It is convenient to configure an IP address on the VIOS because the VIOS can be reached by TCP/IP. For example, you can perform dynamic LPAR operations or enable remote login by configuring an IP address directly on the SEA device, but it can also be defined on an extra VEA in the VIOS that carries the IP address. Doing so leaves the SEA without the IP address, which allows for maintenance on the SEA without losing IP connectivity if SEA failover is configured. Neither approach has a remarkable impact on Ethernet performance.

SEA availability

PowerVM offers a range of configurations to keep the services availability. The following sections present some example scenarios.

Virtual Ethernet redundancy

In a single VIOS configuration, communication to external networks ceases if the VIOS loses connection to the external network. Client partitions experience this disruption if they use the SEA as a means to access the external networks. Communication through the SEA is, for example, suspended when the physical network adapter in the VIOS fails or loses connectivity to the external network due to a switch failure.

Another reason for a failure might be a planned shutdown of the VIOS for maintenance purposes. Communication resumes when the VIOS regains connectivity to the external network. Internal communication between partitions through virtual Ethernet connections continues unaffected while access to the external network is unavailable. Virtual I/O clients do not have to be restarted or otherwise reconfigured to resume communication through the SEA. Similarly, the clients are affected as when unplugging and replugging an uplink of a physical Ethernet switch.

If the temporary failure of communication with external networks is unacceptable, more than a single forwarding instance and some function for failover must be implemented in the VIOS.

Several approaches can be used to achieve HA for shared Ethernet access to external networks. Most commonly used are SEA failover and SEA failover with load sharing, which are described in detail in the following sections.

Other approaches can be used to achieve HA for shared Ethernet access by leveraging configurations that are also used in physical network environments, such as:

- IP Multipathing with DGD or virtual IP addresses (VIPAs) and dynamic routing protocols, such as Open Shortest Path First (OSPF).
- IP Address Takeover (IPAT), with High Availability Cluster Management or Automation Software, such as PowerHA SystemMirror for AIX.

SEA failover

SEA failover offers Ethernet redundancy to the client at the VIOS level. In a SEA failover configuration, two VIOSs have the bridging functions of the SEA. They use a control channel to determine which of them is supplying the Ethernet service to the client. If one SEA loses access to the external network through its physical Ethernet adapter or one VIOS is shut down for maintenance, it automatically fails over to the other VIOS SEA. You also can trigger a manual failover.

The client partition has one VEA that is bridged by two VIOSs. The client partition has no special protocol or software that is configured and uses the VEA as though it was bridged by only one VIOS.

SEA failover supports IEEE 802.1Q VLAN tagging.

As shown in Figure 3-15, both VIOSs attach to the same virtual and physical Ethernet networks and VLANs. Both VEAs of both SEAs have the *access external network* (in a later HMC version, it is “Use this adapter for Ethernet bridging”) flag enabled and a trunk priority (in a later HMC version, it is “priority”) set.

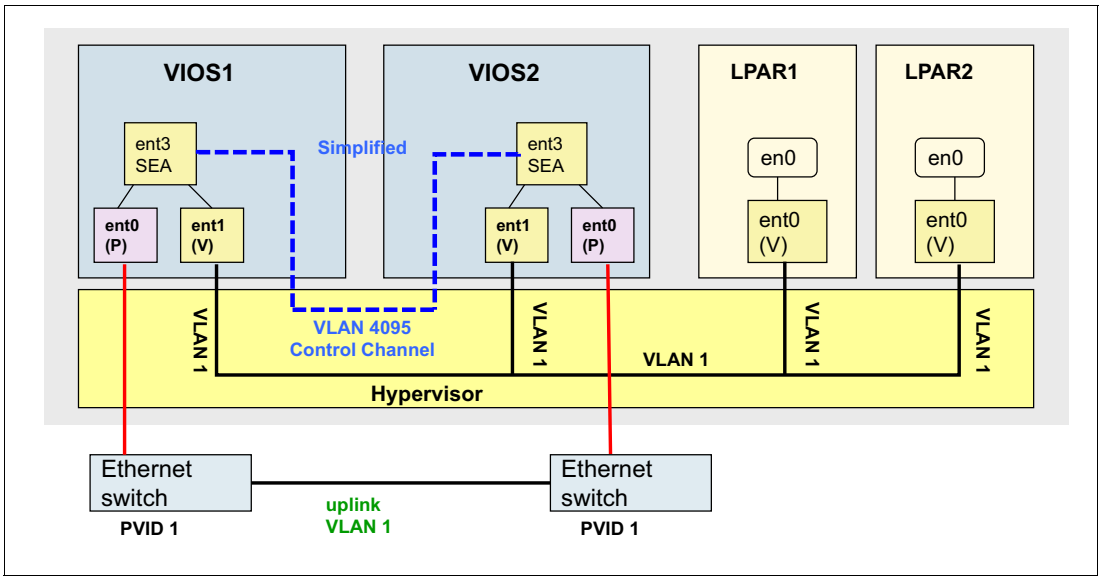


Figure 3-15 Basic SEA failover configuration

An extra virtual Ethernet connection is required as a separate VLAN between the two VIOSs. It must be attached to the SEA as a *control channel*, not as a regular member of the SEA. This VLAN serves as a channel for the exchange of keep-alive or heartbeat messages between the two VIOSs, which controls the failover of the bridging function. When control channel adapters are not configured, VLAN ID 4095 in virtual switch is automatically used for a simplified SEA design. This approach allows SEA partners to heartbeat without dedicated VEAs for control channel.

You must select different priorities for the two SEAs by setting all VEAs of each SEA to that priority value. The priority value defines which of the two SEAs is the primary (active) and which one is the backup (standby). The lower the *priority value*, the higher the priority, so `priority=1` means the highest priority.

Support: SEA failover configurations are supported only on dual-VIOS configurations.

Some types of network failures might not trigger a failover of the SEA because keepalive messages are only sent over the control channel. No keepalive messages are sent over other SEA networks, especially not over the external network. The SEA failover feature can be configured to periodically check the reachability of a specific IP address. The SEA periodically pings this IP address to detect some other network failures. This approach is similar to the IP address ping function that can be configured with NIB.

Important: To use this periodic reachability test, the SEAs must have network interfaces, with IP addresses that are associated. These IP addresses must be unique, and you must use different IP addresses on the two SEAs.

Here are the four cases that initiate a SEA failover:

- The standby SEA detects that keepalive messages from the active SEA are no longer received over the control channel.
- The active SEA detects that a loss of the physical link is reported by the physical Ethernet adapter's device driver.
- On the VIOS with the active SEA, a manual failover can be initiated by setting the active SEA to standby mode.
- The active SEA detects that it cannot ping a specific IP address anymore.

An end of the keepalive messages occurs when the VIOS with the primary SEA is shut down or halted, stops responding, or is deactivated from the HMC.

Important: You might experience up to a 30-second failover delay when SEA failover is used. The behavior depends on the network switch and the spanning tree settings. Any of the following three hints can help in reducing this delay to a minimum:

- For all AIX client partitions, set up DGD on the default route:
  a. Set up DGD on the default route:

     ```
     # route change default -active_dgd
     ```

  b. Add the command **`route change default -active_dgd`** to the `/etc/rc.tcpip` file to make this change permanent.
  c. Set interval between pings of a gateway by DGD to 2 seconds (default is 5 seconds; setting this parameter to 1 or 2 seconds allows faster recovery):

     ```
     # no -p -o dgd_ping_time=2
     ```

- On the network switch, enable Rapid Spanning-Tree (RSTP) or PortFast while legacy Spanning Tree is on, or disable Spanning Tree.
- On the network switch, set the channel group for your ports to **Active** if they are currently set to **Passive**.

Figure 3-16 shows an alternative setup where the IP address of the VIOSs is configured on a separate physical Ethernet adapter.

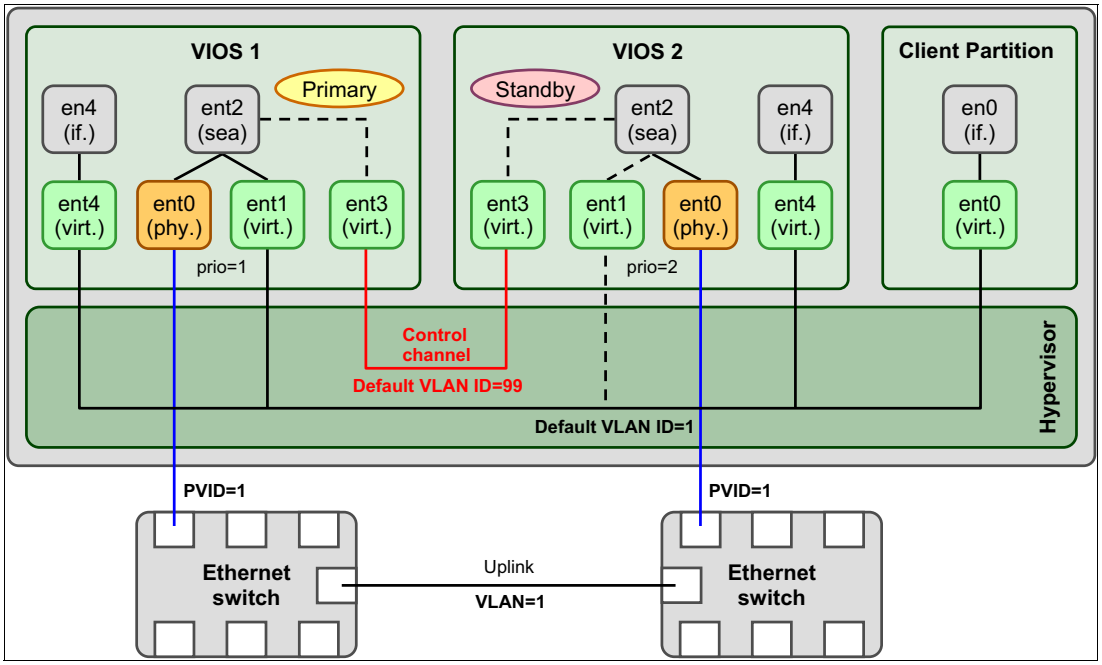


Figure 3-16 Alternative configuration for SEA failover

Network Interface Backup in the client partition

NIB in the client partition can be used to achieve network redundancy when two Virtual I/O Severs (VIOSs) are used. An Etherchannel with only one primary adapter and one backup adapter is said to be operating in NIB mode.

Figure 3-17 shows an NIB setup for an AIX client partition. The client partition uses two VEAs to create an Etherchannel that consists of one primary adapter and one backup adapter. The interface is defined on the Etherchannel. If the primary adapter becomes unavailable, the NIB switches to the backup adapter.

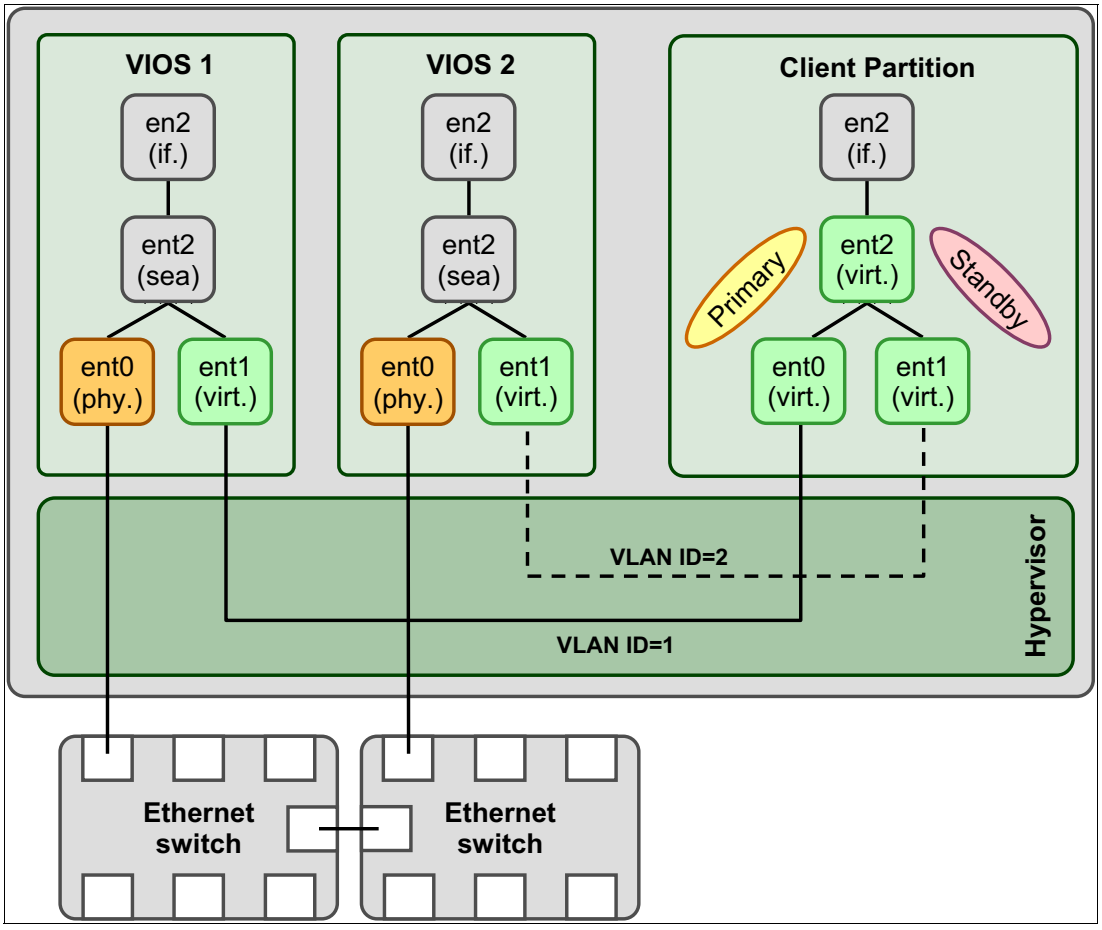


Figure 3-17 Network redundancy by using two Virtual I/O Servers and NIB

An LA of more than one active VEA is not supported. Only one primary VEA plus one backup VEA are supported. To increase the bandwidth of a VEA, LA must be done on the VIOS.

When NIB is configured in a client partition, each VEA must be configured on a different VLAN.

Important: When NIB is used with VEAs on AIX, you must use the ping-to-address feature to detect network failures. The reason is that there is no hardware link failure for VEAs to trigger a failover to the other adapter.

For IBM i, an equivalent solution to NIB can be implemented by using VIPA failover with a virtual-to-VEA failover script. The same solution can be implemented on Linux VMs by using Ethernet connection bonding.

SEA failover with load sharing

The VIOS provides a load-sharing function to enable the usage of the bandwidth of the backup SEA.

In a SEA failover configuration, the backup SEA is in standby mode, and is used only when the primary SEA fails. The bandwidth of the backup SEA is not used in normal operation.

Figure 3-15 on page 125 shows a basic SEA failover configuration. All network packets of all Virtual I/O clients are bridged by the primary VIOS.

A SEA failover with load sharing effectively uses the backup SEA bandwidth, as shown in Figure 3-18. In this example, network packets of for VLANs 12 and 14 are bridged by VIOS2, where VLANs 11 and 13 are bridged by VIOS1.

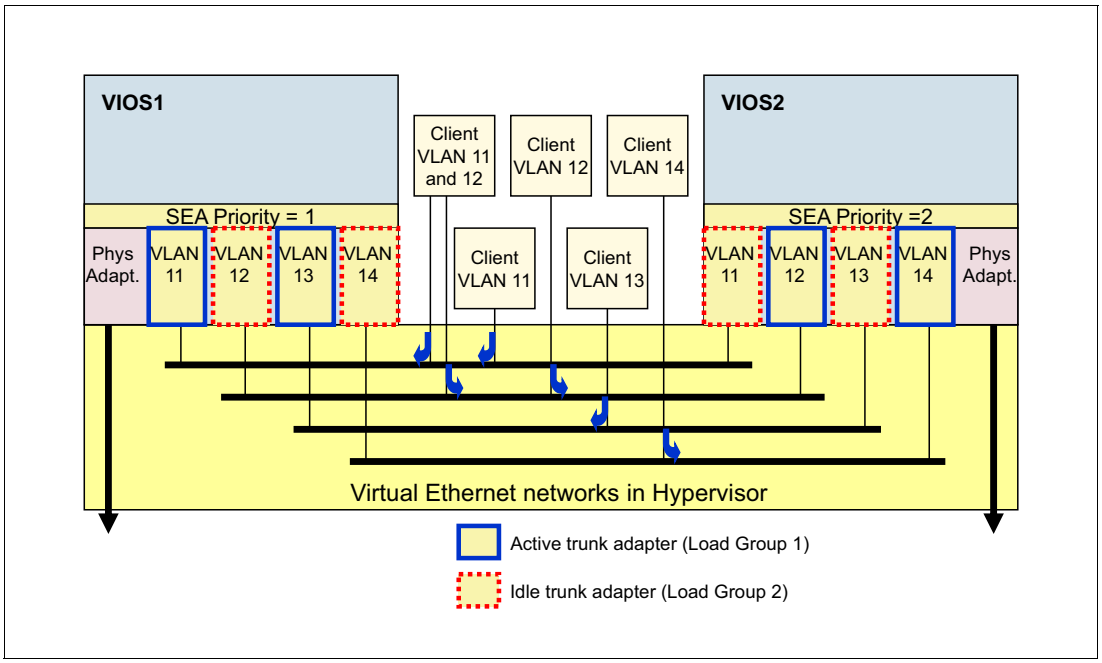


Figure 3-18 SEA failover with load-sharing

Prerequisites for SEA failover with load sharing are as follows:

- Both primary and backup VIOSs are at Version 2.2.1.0 or later.
- Two or more trunk adapters are configured for the primary and backup SEA pairs.
- Load-sharing mode must be enabled on both the primary and backup SEA pair.
- The VLAN definitions of the trunk adapters are identical between the primary and backup SEA pair.

Important: You must set the same priority to all trunk adapters under one SEA. The primary and backup priority definitions are set at the SEA level, not at the trunk adapters level.

Using link aggregation on the Virtual I/O Server

LA is a network port aggregation technology that allows several Ethernet adapters to be aggregated together to form a single pseudo-Ethernet adapter. This technology can be used on the VIOS to increase the bandwidth compared to when a single network adapter is used. It also avoids bottlenecks when one network adapter is shared among many client partitions.

The main benefit of an LA is that it has the network bandwidth of all its adapters in a single network presence. If an adapter fails, the packets are automatically sent to the next available adapter without disruption to existing user connections. The adapter is automatically returned to service on the LA when it recovers. Thus, LA also provides some degree of increased availability. A link or adapter failure leads to a performance degradation, but not a disruption.

Depending on the manufacturer, LA is not a complete HA networking solution because all the aggregated links must connect to the same switch. By using a backup adapter, you can add a single extra link to the LA, which is connected to a different Ethernet switch with the same VLAN. This single link is used only as a backup.

As an example for LA, ent0 and ent1 can be aggregated to ent2. The system considers these aggregated adapters as one adapter. Then, interface en2 is configured with an IP address. Therefore, IP is configured as on any other Ethernet adapter. In addition, all adapters in the LA are given the same hardware (MAC) address so that they are treated by remote systems as though they were one adapter.

Two variants of LA are supported:

- Cisco Etherchannel
- IEEE 802.3ad Link Aggregation

Although Etherchannel is a Cisco-specific implementation of adapter aggregation, LA follows the IEEE 802.3ad standard. Table 3-5 shows the main differences between Etherchannel and LA.

Table 3-5 Main differences between Etherchannel and Link Aggregation

Cisco Etherchannel	IEEE 802.3ad Link Aggregation
Cisco-specific.	Open standard.
Requires switch configuration.	Little, if any, configuration of the switch is required to form aggregation. Some initial setup of the switch might be required.
Supports different packet distribution modes.	Supports only standard distribution mode.

Using IEEE 802.3ad Link Aggregation allows for the use of Ethernet switches, which support the IEEE 802.3ad standard but might not support Etherchannel. The benefit of Etherchannel is the support of different packet distribution modes. This support means that it is possible to influence the load-balancing of the aggregated adapters. In the remainder of this publication, we use *LA* where possible because that is considered a more universally understood term.

Note: When IEEE 802.3ad Link Aggregation is used, ensure that your Ethernet switch hardware supports the IEEE 802.3ad standard. In VIOS, configuring an Ethernet interface to use the 802.3ad mode requires that the Ethernet switch ports also are configured in IEEE 802.3ad mode.

Figure 3-19 shows the aggregation of two plus one adapters to a single pseudo-Ethernet device, including a backup feature.

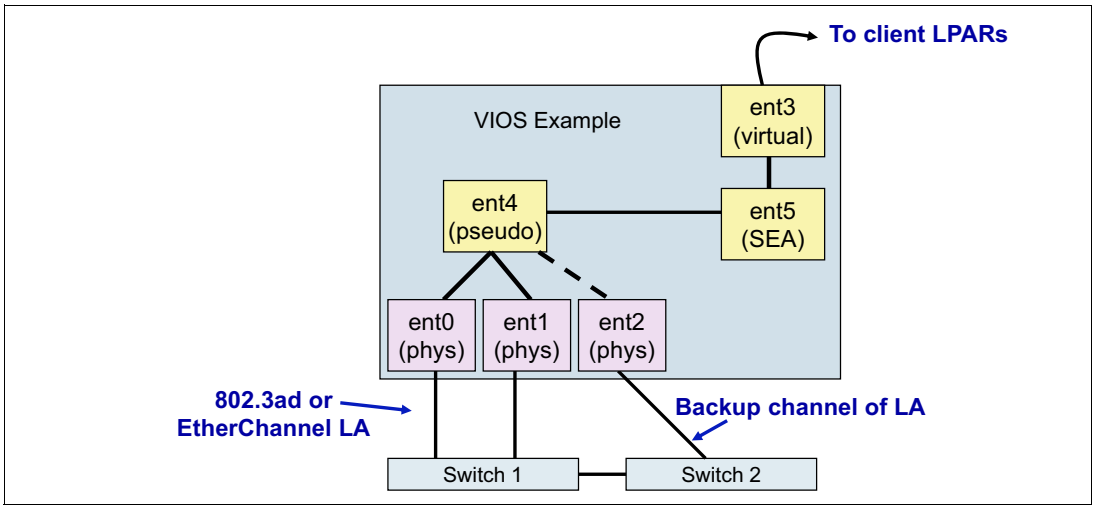


Figure 3-19 Link aggregation (Etherchannel) on the Virtual I/O Server

The Ethernet adapters ent0 and ent1 are aggregated for bandwidth and must be connected to the same Ethernet switch, and ent2 connects to a different switch. ent2 is used only for backup, for example, if the main Ethernet switch fails. The adapters ent0 and ent1 are exclusively accessible through the pseudo-Ethernet adapter ent5 and its interface en5. You cannot, for example, attach a network interface en0 to ent0 if ent0 is a member of an Etherchannel or LA.

Support: A LA or Etherchannel of VEAs is not supported, but you can use the NIB feature of LA with VEAs.

A LA with only one primary Ethernet adapter and one backup adapter is operating in NIB.

For examples and scenarios of networking configurations for the VIOS LPAR and the client LPARs, see Scenarios: Configuring the Virtual I/O Server, found at:

`https://www.ibm.com/docs/en/power10/9080-HEX?topic=server-scenarios`

SEA quality of service

The SEA can enforce quality of service (QoS) based on the IEEE 802.1q standard. This section explains how QoS works for SEA and how it can be configured.

SEA QoS provides a means where the VLAN tagged egress traffic is prioritized among seven priority queues. However, QoS comes into play only when contention is present.

Each SEA instance has some threads (currently seven) for multiprocessing. Each thread has nine queues to take care of network jobs. Each queue takes care of jobs at a different priority level. One queue is kept aside and used when QoS is disabled.

Important: QoS works only for tagged packets, that is, all packets that emanate from the VLAN pseudo-device of the virtual I/O client. Therefore, because virtual Ethernet does not tag packets, its network traffic cannot be prioritized. The packets are placed in queue 0, which is the default queue at priority level 1.

Each thread independently follows the same algorithm to determine from which queue to send a packet. A thread sleeps when no packets are available on any of the nine queues.

Note the following points:

- If QoS is enabled, SEA checks the priority value of all tagged packets and puts that packet in the corresponding queue.
- If QoS is *not* enabled, then regardless of whether the packet is tagged or untagged, SEA ignores the priority value and places all packets in the disabled queue. This approach ensures that the packets that are enqueued while QoS is disabled are not sent out of order when QoS is enabled.

When QoS is enabled, two algorithms are available to schedule jobs: strict mode and loose mode.

Strict mode

In strict mode, all packets from higher priority queues are sent before any packets from a lower priority queue. The SEA examines the highest priority queue for any packets to send out. If any packets are available to send, the SEA sends that packet. If no packets are available to send in a higher priority queue, the SEA checks the next highest priority queue for any packets to send out.

After a packet from the highest priority queue with packets is sent out, the SEA starts the algorithm over again. This approach allows for high priorities to be serviced before the lower priority queues.

Loose mode

It is possible, in strict mode, that lower priority packets are never serviced if higher priorities packets always are present. To address this issue, the loose mode algorithm was devised.

With loose mode, if the number of bytes that is allowed already was sent out from one priority queue, then the SEA checks all lower priorities at least once for packets to send before packets from the higher priority are sent again.

When packets are initially sent out, SEA checks its highest priority queue. It continues to send packets out from the highest priority queue until either the queue is empty, or the cap is reached. After either of those two conditions are met, SEA moves on to service the next priority queue. It continues by using the same algorithm until either of the two conditions are met in that queue. At that point, it moves on to the next priority queue. On a fully saturated network, this process allocates certain percentages of bandwidth to each priority. The caps for each priority are distinct and nonconfigurable.

A cap is placed on each priority level so that after a number of bytes is sent for each priority level, the following level is serviced. This method ensures that all packets are eventually sent. More important traffic is given less bandwidth with this mode than with strict mode. However, the caps in loose mode are such that more bytes are sent for the more important traffic, so it still gets more bandwidth than less important traffic. Set loose mode by using this command:

```
chdev -dev  -attr qos_mode=loose
```

You can dynamically configure the QoS priority of a VEA of a running LPAR by using the HMC. You can prioritize the LPAR network traffic by specifying the value of IEEE 802.1Q priority level for each VEA.

SEA performance considerations

When virtual networking is used, some performance implications must be considered. Therefore, networking configurations are site-specific. For this reason, no guaranteed rules for performance tuning exist.

The following considerations apply to VEA and SEA:

- The usage of VEA in a partition does not increase its CPU requirement. However, high levels of network traffic within a partition increase CPU utilization. This behavior is not specific to virtual networking configurations.
- The usage of SEA in a VIOS increases the CPU utilization of the partition due to the bridging function of the SEA.
- Keep the threading option enabled (default) on the SEA when the VIOS also is hosting virtual storage (vSCSI or NPIV).
- SEA configurations that use high-speed physical adapters can be demanding on CPU resources within the VIOS. Ensure that you assign sufficient CPU capacity to the VIOS.
- To reduce CPU processing overhead for TCP workloads on the VIOS and client partitions and to better use the wire speed of high-speed Ethernet adapters:
  - Enable large send offload (LSO) on the client partition's interface (on the VIOS it is enabled by default).
  - Enable large receive offload on the SEA of the VIOS.

Notes:

- For IBM i, large receive offload is supported by IBM i 7.1 TR5 and later.
- Large receive offload by default is disabled on the VIOS's *SEA* to eliminate incompatibility with older Linux distributions. Consider enabling large receive on SEA when a supported Linux distribution is used.

- Consider the usage of jumbo frames and increasing the MTU to 9000 bytes if possible when high-speed adapters are used. Jumbo frames enable higher throughput for fewer CPU cycles. However, the external network also must be configured to support the larger frame size.

For more information about tuning network performance throughput, see *IBM PowerVM Virtualization Managing and Monitoring*, SG24-7590.

SEA network requirements planning

For network planning guidance for SEA network design with high-speed adapters, see Network requirements, found at:

`https://www.ibm.com/docs/en/power10/9080-HEX?topic=adapters-network-requirements`

The attributes and performance characteristics of various types of Ethernet adapters help you select which adapters to use in your environment. For more information, see Adapter selection, found at:

`https://www.ibm.com/docs/en/power10/9080-HEX?topic=adapters-adapter-selection`

Processor allocation guidelines exist for both dedicated processor LPARs and shared processor LPARs. Because Ethernet running MTU size of 1500 bytes consumes more processor cycles than Ethernet running jumbo frames (MTU 9000), the guidelines are different for each situation. In general, the processor utilization for large packet workloads on jumbo frames is approximately half that required for MTU 1500. For more information, see Processor allocation, found at:

https://www.ibm.com/docs/en/power10/9080-HEX?topic=adapters-processor-allocation

In general, 512 MB of memory per LPAR is sufficient for most configurations. Enough memory must be allocated for the VIOS data structures. Ethernet adapters and virtual devices use dedicated receive buffers. These buffers are used to store the incoming packets, which are then sent over the outgoing device.

A physical Ethernet adapter typically uses 4 MB for MTU 1500 or 16 MB for MTU 9000 for dedicated receive buffers for gigabit Ethernet. Other Ethernet adapters are similar. Virtual Ethernet typically uses 6 MB for dedicated receive buffers. However, this number can vary based on workload. Each instance of a physical or virtual Ethernet needs memory for this number of buffers. In addition, the system has a mbuf buffer pool per processor that is used if extra buffers are needed. These mbufs typically occupy 40 MB. For more information, see Planning for Shared Ethernet Adapters, found at:

https://www.ibm.com/docs/en/power10/9080-HEX?topic=planning-shared-ethernet-adapters

Enabling largesend and jumbo_frames

IBM AIX allows you to transmit large packets and frames through a network. To send a large data chunk over the network, TCP breaks it down into multiple segments, which requires multiple calls down the stack and results in higher processor utilization on the host processor. You can address the issue by using the TCP LSO option, which allows the AIX TCP layer to build a TCP message that is up to 64 KB long.

For example, without the TCP LSO option, sending 64 KB of data takes 44 calls down the stack by using 1500-byte Ethernet frames. With the TCP LSO option enabled, the TCP option can send up to 64 KB of data to the network interface card (NIC) in a single transmit-receive call. In a real-time scenario, the required number of processor cycles is controlled by the application and depends on the speed of the physical network. With faster networks, the usage of LSO reduces the host processor utilization and increases throughput.

A jumbo frame is an Ethernet frame with a payload greater than the standard MTU of 1,500 bytes and can be as large as 9,000 bytes. It has the potential to reduce processor usage.

TCP LSO and jumbo_frames in AIX are independent of each other. They can be used together or in isolation. For more information about how to enable largesend and jumbo_frames, see Enabling largesend and jumbo_frames in IBM AIX to reduce processor usage, found at:

https://developer.ibm.com/articles/au-aix-largesend-jumboframes/

3.6.5 SR-IOV planning

SR-IOV is described in 2.4.3, “Single-root I/O virtualization” on page 47.

An SR-IOV architecture defines virtual replicas of PCI functions that are known as *virtual functions* (VFs). An LPAR can connect directly to an SR-IOV adapter VF without going through a virtual intermediary (VI) such as a PHYP or VIOS. This ability provides for a low latency and lower CPU utilization alternative by avoiding a VI.

An SR-IOV-capable adapter might be assigned to an LPAR in dedicated mode or enabled for shared mode. The management console provides an interface to enable SR-IOV shared mode.

An SR-IOV-capable adapter in shared mode is assigned to the hypervisor for management of the adapter and provisioning of adapter resources to LPARs. With the management console, along with the hypervisor, you can manage the adapter's physical Ethernet ports and logical ports (LPs).

To connect an LPAR to an SR-IOV Ethernet adapter VF, create an SR-IOV Ethernet LP for the LPAR. When you create an Ethernet LP for a partition, select the adapter physical Ethernet port to connect to the LPAR and specify the resource requirements for the LP. Each LPAR can have one or more LPs from each SR-IOV adapter in shared mode. The number of LPs for all configured LPARs cannot exceed the adapter LP limit.

Note: An SR-IOV adapter does not support LPM unless the VF is assigned to a SEA or used together with vNIC.

For an SR-IOV adapter in shared mode, the physical port switch mode can be configured in Virtual Ethernet Bridge (VEB) mode, which is the default setting, or Virtual Ethernet Port Aggregator (VEPA) mode. If the switch is configured in VEB mode, the traffic between the LPs is not visible to the external switch. If the switch is configured in VEPA mode, the traffic between LPs must be routed back to the physical port by the external switch. Before you enable the physical port switch in VEPA mode, ensure that the switch that is attached to the physical port is supported and enabled for reflective relay.

When bridging between VEAs and a physical Ethernet adapter, an SR-IOV Ethernet LP might be used as the physical Ethernet adapter to access the outside network. When an LP is configured as the physical Ethernet adapter for bridging, promiscuous permission must be enabled in the LP. For example, if you create an LP for a VIOS LPAR and the intent is to use the LP as the physical adapter for the SEA, you must select the promiscuous permission for the LP.

Configuration requirements

Consider the following configuration requirements when an Ethernet LP is used as the physical Ethernet device for SEA bridging:

- If diverting all network traffic to flow through an external switch is required, consider the following requirements:
  - The hypervisor virtual switch must be set to the VEPA switching mode, and the SR-IOV Ethernet adapter physical port switch mode must be set to the VEPA switching mode.
  - In addition, the LP is the only LP that is configured for the physical port.
- When you create an Ethernet LP, you can specify a capacity value. The capacity value specifies the required capacity of the LP as a percentage of the capability of the physical port. The capacity value determines the number of resources that are assigned to the LP from the physical port. The assigned resources determine the minimum capability of the LP. Physical port resources that are not used by other LPs might be temporarily used by the LP when the LP exceeds its assigned resources to allow extra capability. System or network limitations can influence the amount of throughput an LP can achieve. The maximum capacity that can be assigned to an LP is 100%. The sum of the capacity values for all the configured LPs on a physical port must be less than or equal to 100%. To minimize the configuration effort while more LPs are added, you might want to reserve physical port capacity for extra LPs.

- When an Ethernet LP is used as a physical adapter for bridging VEAs, the parameter values such as the number of client virtual adapters and expected throughput must be considered when a capacity value is chosen.
- The Ethernet LPs allow the LP to run diagnostics on the adapter and physical port. Select this permission only while the diagnostics are run by using the LP.

Verifying that the server supports single-root I/O virtualization

Before you enable SR-IOV shared mode for an SR-IOV-capable adapter, verify that the server supports the SR-IOV feature by using the HMC.

To verify that the server supports SR-IOV, complete the following steps:

1. In the navigation pane, click **Resources**.
2. Click **All Systems**. The **All Systems** window opens.
3. In the work pane, select the system and select **Actions** → **View System Properties**. The **Properties** window opens.
4. Click **Licensed Capabilities**. The Licensed Capabilities window lists the features that are supported by the server.
5. In the Licensed Capabilities window, verify the list of features that are displayed:
   - If **SR-IOV Capable** is marked by the check mark icon, which represents the availability of a feature in the HMC icon, the SR-IOV adapter can be configured in the shared mode and shared by multiple LPARs.
   - If **SR-IOV Capable** is marked by the -- icon, which represents the nonavailability of a feature in the HMC icon, the SR-IOV adapter can be configured in the shared mode, but can be used by only one LPAR.
   - If **SR-IOV Capable** is not displayed, the server does not support the SR-IOV feature.
6. Click **OK**.

Verifying the logical port limit and the owner of the SR-IOV adapter

You can view the LP limit and the owner of the SR-IOV adapter by using the HMC. To view the LP limit and the owner of the SR-IOV adapter, complete the following steps:

1. In the navigation pane, click **Resources**.
2. Click **All Systems**. The **All Systems** window opens.
3. In the work pane, select the system and select **Actions** → **View System Properties**. The **Properties** window opens.
4. Click **Licensed Capabilities**. The Licensed Capabilities window lists the features that are supported by the server.
5. In the Properties area, click the **Processor, Memory, I/O** tab. In the Physical I/O Adapters area, the table displays the SR-IOV capable (Logical Port Limit) and the Owner details about the SR-IOV adapter.

- The SR-IOV capable (Logical Port Limit) column displays whether the slot or the adapter is SR-IOV capable, and the maximum number of LPs that this slot or the adapter can support. If the slot or the adapter is SR-IOV-capable but is assigned to a partition, the SR-IOV capable (Logical Port Limit) column indicates that the slot or the adapter is in the dedicated mode.

- The Owner column displays the name of the current owner of the physical I/O. The value of this column can be any of the following values:
  - When an SR-IOV adapter is in shared mode, a hypervisor is displayed in this column.
  - When an SR-IOV adapter is in dedicated mode, Unassigned is displayed when the adapter is not assigned to any partition as a dedicated physical I/O.
  - When an SR-IOV adapter is in dedicated mode, the LPAR name is displayed when the adapter is assigned to any LPAR as a dedicated physical I/O.

3.6.6 SR-IOV with vNIC planning

To configure a vNIC client, an adapter must be configured in SR-IOV shared mode before a vNIC client is configured. In addition, LPs must be available and physical port capacity must be available. So, the total of activated LP capacity values for the physical port must be less than 100%.

Some limits also apply to the number of vNIC adapters for a partition. FW840.10 allows 10 client vNIC adapters per partition.

Partitions that are configured with vNIC adapters are compatible with LPM and SRR technologies.

Some minimum code levels that support vNIC for each operating system are required. For more information about the exact requirements for your target platform, see PowerVM vNIC and vNIC Failover FAQs, found at:

https://community.ibm.com/HigherLogic/System/DownloadDocumentFile.ashx?DocumentFileKey=96088528-4283-8b61-38b0-a39c9ed990c7&forceDialog=0

LA is supported if a vNIC client has a single backing device. A vNIC client with multiple backing devices (vNIC failover) in combination with LA technologies such as IEEE802.3ad/802.1ax (LACP), AIX NIB, or Linux bonding active backup mode is not supported. SR-IOV LA limitations apply to client vNIC adapters.

vNIC failover considerations

vNIC failover allows a vNIC client to be configured with up to six backing devices. One backing device is active while the others are inactive standby devices. If the hypervisor detects that the active backing device is no longer operational, a failover is initiated to the most favored (lowest Failover Priority value) operational backing device.

Some minimum code levels are required for vNIC failover. In general, HMC, system firmware, and operating systems with support for Power10 processor-based servers include support for vNIC failover.

Backing devices can be dynamically added and removed to a vNIC client.

When backing devices are designed, consider combining separate SR-IOV adapters with different VIOS for the same partitions for redundancy purposes. Example backing devices for a single client vNIC adapter are as follows:

- vNIC Backing Device 1: SR-IOV Adapter 1 that uses VIOS1
- vNIC Backing Device 2: SR-IOV Adapter 2 that uses VIOS2

A comparison of network virtualization technologies can be found in Table 3-6.

Table 3-6 A comparison of network virtualization technologies

Technology	LPM support	QoS	Direct-access performance	Redundancy options	Server-side redundancy	Requires VIOS
SR-IOV	No[a]	Yes	Yes	Yes[b]	No	No
vNIC	Yes	Yes	No[c]	Yes[b]	vNIC Failover	Yes
SEA or virtual Ethernet	Yes	Yes[d]	No	Yes	SEA Failover	Yes
Hybrid Network Virtualization (HNV)	Yes	Yes	Yes	Yes	No	No[e]

a. SR-IOV optionally can be used as the backing device of SEA in VIOS to use higher-level virtualization functions like LPM. However, the client partition does not receive the performance or QoS benefit.
b. Some limitations apply. For more information, see FAQs on LA, found at https://community.ibm.com/community/user/power/viewdocument/sr-iov-vnic-and-hnv-information.
c. Generally, provides better performance and requires fewer system resources compared to SEA or virtual Ethernet.
d. SR-IOV has a superior QoS capability compared to SEA.
e. VIOS is not required during regular operations. However, VIOS is used to host the backup (vNIC) adapter during LPM operations.

3.7 Further considerations

For more information about best practices, recommendations, and special considerations, see the following resources:

- *IBM Power Virtualization Best Practices Guide,* found at:

  https://www.ibm.com/downloads/cas/JVGZA8RW

- *Power10 Performance Quick Start Guides*, found at:

  https://www.ibm.com/support/pages/system/files/inline-files/Power10_Performance_Quick_Start_Guides.pdf

- Power10 Performance Best Practices - A brief checklist, found at:

  https://www.ibm.com/support/pages/system/files/inline-files/power10_performance_best_practices.pdfD

4

Implementing IBM PowerVM

This chapter describes the implementation and configuration of PowerVM features. The information in this chapter helps new PowerVM users to get started with implementation. The PowerVM features that are described in this chapter were introduced in Chapter 1, “IBM PowerVM overview” on page 1and Chapter 2, “IBM PowerVM features in details” on page 31. The planning aspects are covered in Chapter 3, “Planning for IBM PowerVM” on page 77.

This chapter covers the following topics:

- Adding the managed system to the Hardware Management Console
- Creating, installing, and configuring a Virtual I/O Server logical partition
- Network configuration
- Creating and installing a client LPAR
- VIOS security implementation
- Shared processor pools
- Active Memory Expansion implementation
- Active Memory Mirroring implementation
- PowerVC Implementation

4.1 Adding the managed system to the Hardware Management Console

This section describes how to configure and connect a new Power10 processor-based enterprise Baseboard Management Controller (eBMC)-managed system to the Hardware Management Console (HMC).

For a demonstration about how to set up a new Power10 processor-based eBMC-based system where the initial IP address is provided by the HMC Dynamic Host Configuration Protocol (DHCP) server, see Configuring a new Power10 eBMC system with a DHCP address that is provided by the HMC, found at:

https://mediacenter.ibm.com/media/Configuring+a+new+POWER10+eBMC+system+with+a+DHCP+address+provided+by+the+HMC/1_lznm2f7r

4.1.1 eBMC and Virtualization Management Interface configuration

When a new system is added to a DHCP-enabled HMC, the system receives an IP address from the HMC upon request.

One unique feature about eBMC is that when a system is connected to a DHCP-enabled HMC, the Serial Number column on HMC does not show the serial number of the system. It shows the IP address that was assigned by DHCP without the periods, as shown in Figure 4-1. It is not the serial number of the machine, but the IP address that was given.

Figure 4-1 also shows that authentication is pending and that a password must be entered.

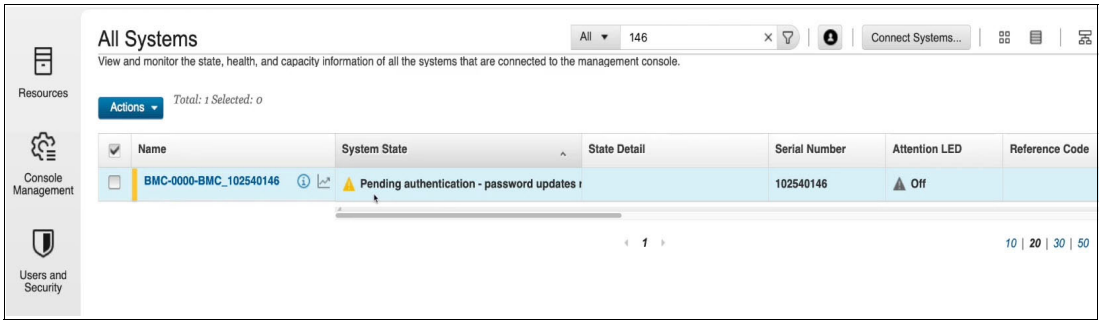


Figure 4-1 HMC eBMC connection

To update the password, select the system and from the **Actions** menu, select **Update System Password** as shown in Figure 4-2 on page 141. This procedure is the same for Flexible Service Processor (FSP)-based systems. The difference is that the HMC access password and the Advanced System Management Interface (ASMI) Admin password are the same on eBMC-based systems. This password requires higher complexity than on FSP-based systems.

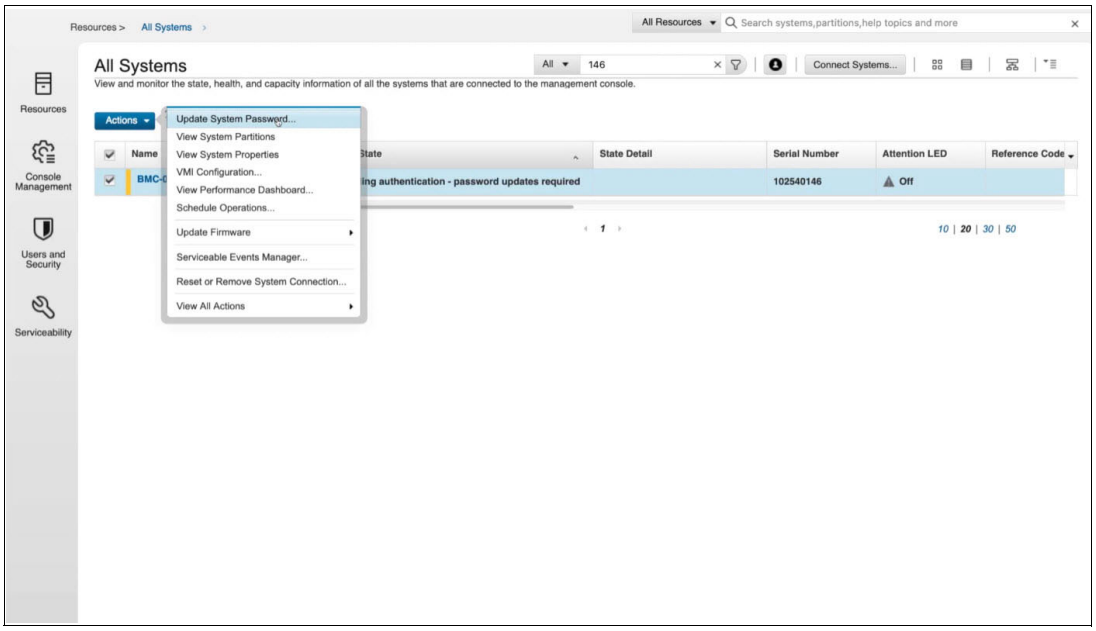


Figure 4-2 Update System Password

After the password is updated, configure Virtualization Management Interface (VMI) (Figure 4-3).

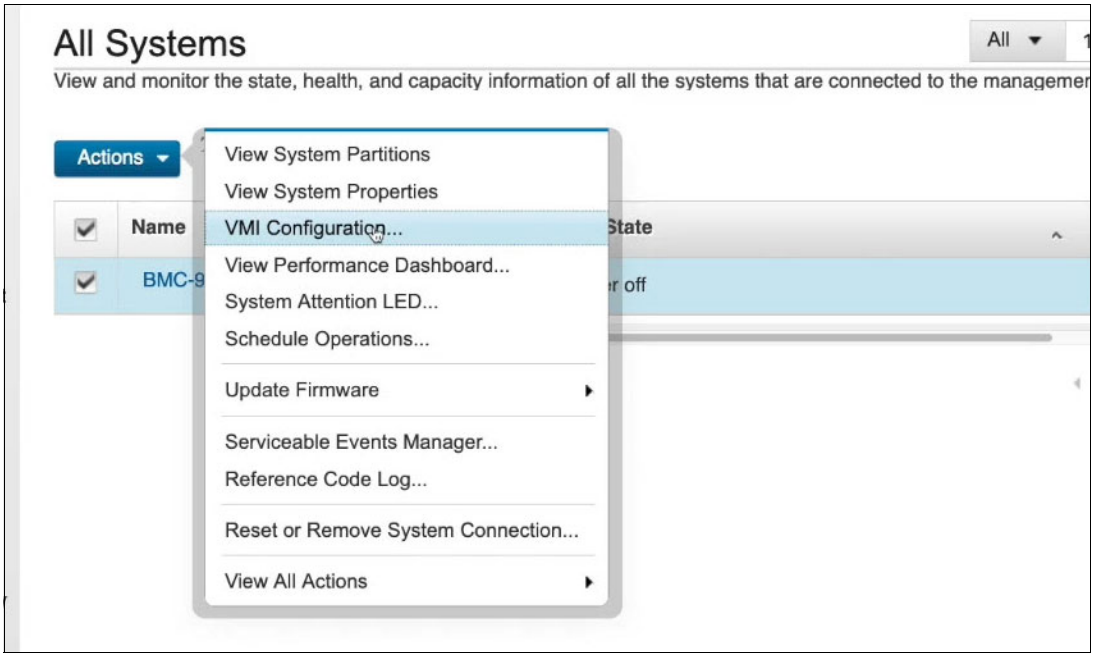


Figure 4-3 VMI configuration

VMI supports both static and DHCP IP configurations. You can use the HMC command-line interface (CLI), GUI, or REST API to view and configure the VMI IP address.

For more information, see VMI configuration for eBMC-based managed systems, found at:

https://www.ibm.com/docs/en/power10/000V-HMC?topic=operations-vmi-configuration-ebmc-based-managed-systems

4.2 Creating, installing, and configuring a Virtual I/O Server logical partition

This section describes the steps that are required to create a Virtual I/O Server (VIOS) logical partition (LPAR), and the required resources to assign to the VIOS LPAR. The installation methods that are available for the VIOS operating system also are described.

4.2.1 HMC versus PowerVM NovaLink managed environment

Before we explain the implementation of a VIOS LPAR, we must describe the available virtualization management consoles for Power servers:

- HMC
- PowerVM NovaLink

Both HMC and PowerVM NovaLink can configure and control managed systems (Power servers), which includes the creation of VIOS and client LPARs. However, some differences exist between the two products in their features and architecture. For example, PowerVM NovaLink is installed and runs directly on a thin Linux LPAR in the managed system. On each managed system in the environment, a thin Linux partition exists to run the compute processes for its own managed system. However, the HMC does not consume any resources from the systems that it is managing. Also, it can manage and control multiple managed systems and their compute processes. The advantages of PowerVM NovaLink become apparent when it is combined with IBM Power Virtualization Center (PowerVC).

For more information, see Management console comparison, found at:

https://www.ibm.com/docs/en/power10?topic=powervm-management-console-comparison

This publication focuses on the implementation of PowerVM features by using the HMC Enhanced GUI. For more information about PowerVM NovaLink installation, creation, and installation of VIOS, see PowerVM NovaLink, found at:

https://www.ibm.com/docs/en/power10?topic=environment-powervm-novalink

4.2.2 Creating the VIOS LPAR on an HMC-managed environment

This section describes the creation of the VIOS LPAR by using the HMC Enhanced GUI.

System plans: You can use the HMC to create a system plan that is based on an existing managed system configuration. Then, deploy that system plan to other managed systems. For more information about system plans, see System plans on the HMC, found at:

https://www.ibm.com/docs/en/power10?topic=plans-system-hmc

To create a VIOS LPAR in an HMC-managed environment, complete the following steps:

1. On HMC Enhanced GUI, select **Resources** from the left pane. Then, select **All Systems** to view the managed systems. as shown in Figure 4-4.

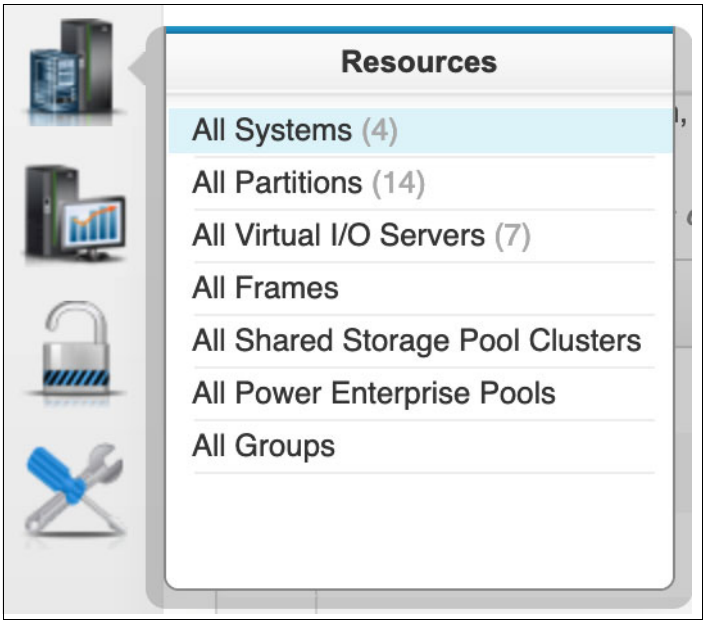


Figure 4-4 Available resources that are managed by HMC.

After selecting **All Systems**, the next window shows managed systems that are connected to the HMC, as shown in Figure 4-5.

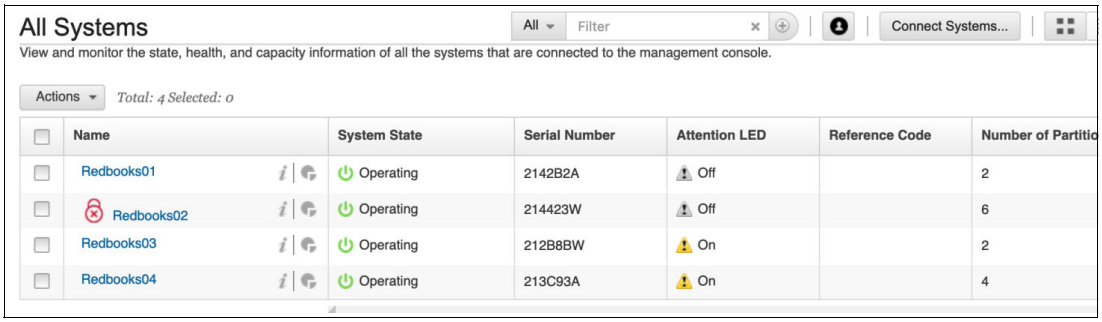


Figure 4-5 Managed systems on the HMC Enhanced GUI

2. Click the name of a managed system to view the managed system window. Then, select **Virtual I/O Servers** from the left pane, and then select **Create Virtual I/O Server**, as shown in Figure 4-6.

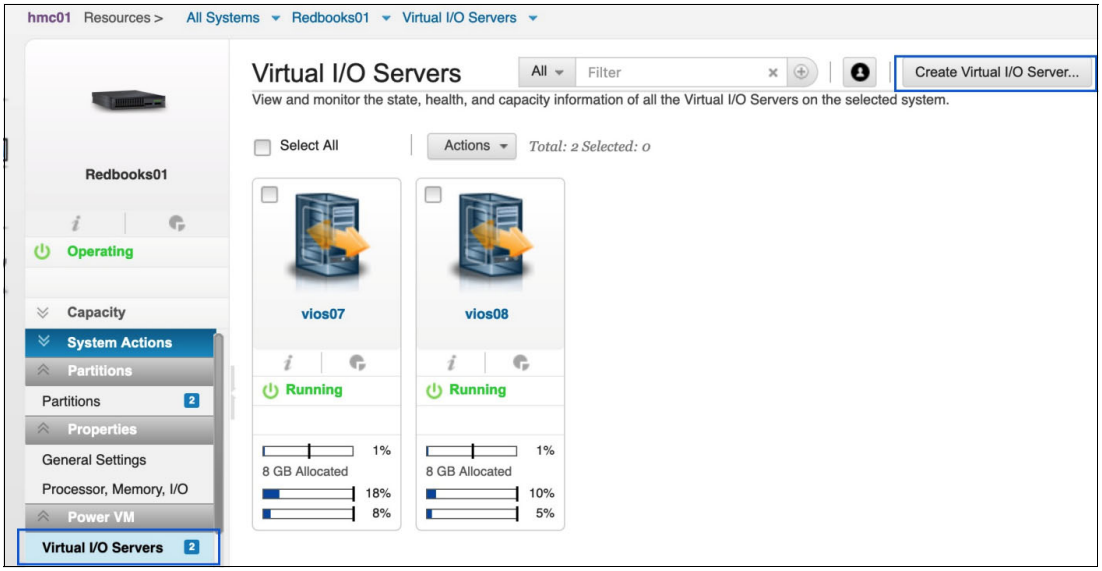


Figure 4-6 Virtual I/O Servers window to start the VIOS LPAR creation wizard

3. Enter the VIOS LPAR name and partition ID. By default, the partition ID shows the next available partition ID number. The Partition ID must be a unique number. Click **Next**.
4. Select whether the processors are part of a shared pool or dedicated for this partition. If shared is selected, the partition will be a micro-partition (Figure 4-7). Click **Next**.

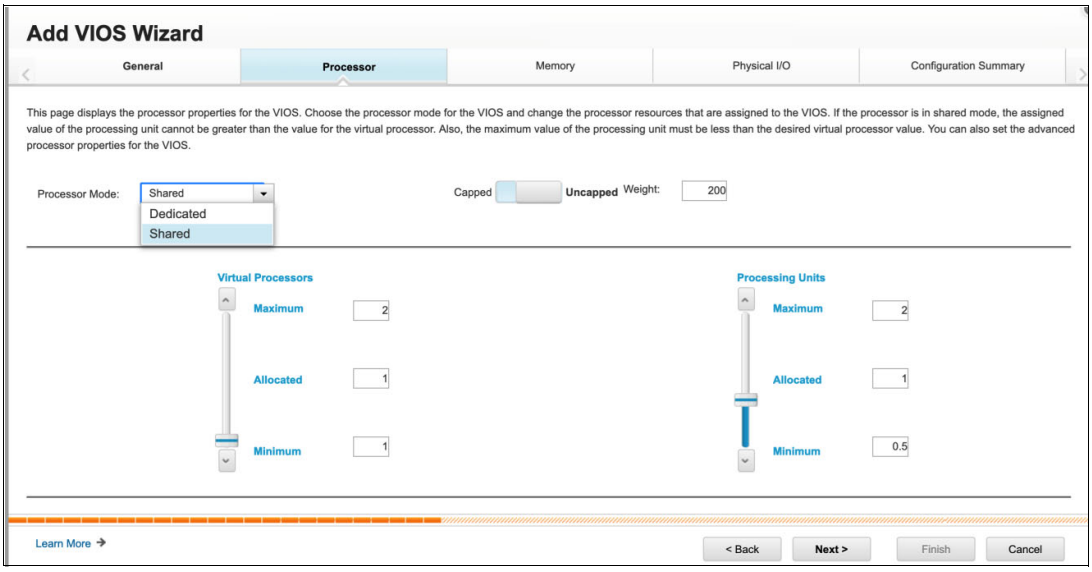


Figure 4-7 HMC Virtual I/O Server processor settings for a micro-partition

Rules: The following rules apply to the processor settings:

- The system tries to allocate the wanted values.
- The partition does not start if the managed system cannot provide the minimum number of processing units.
- You cannot dynamically increase the number of processing units to more than the defined maximum. If you want more processing units, the partition must be stopped, and then reactivated with an updated profile (not only restarted).
- The maximum number of processing units cannot exceed the total Managed System processing units.

5. Choose the memory setting and click **Next** (Figure 4-8).

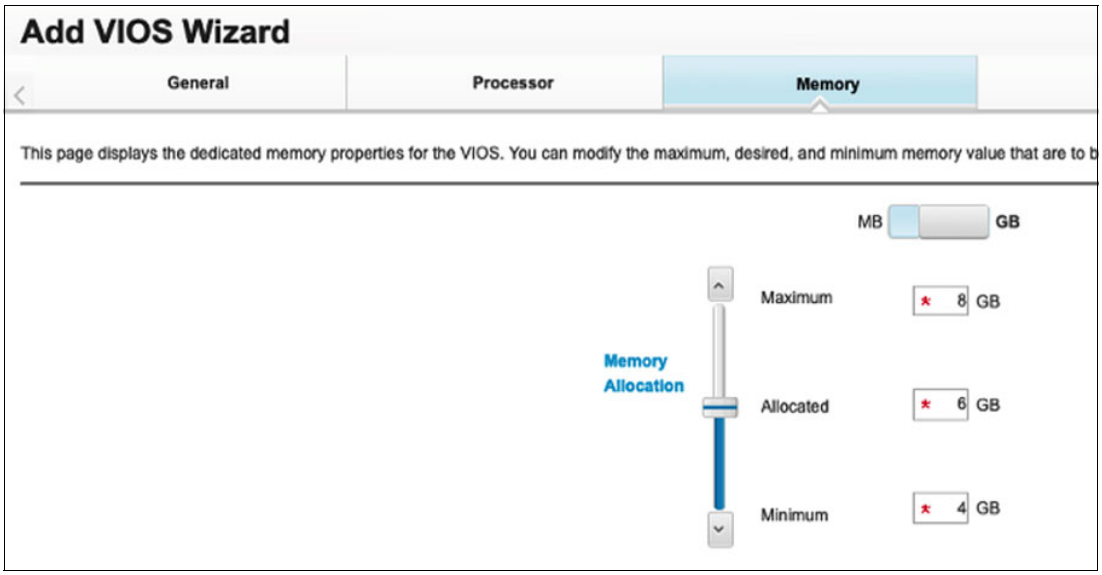


Figure 4-8 HMC Virtual I/O Server memory settings

Rules: The following rules apply to the system that is shown in Figure 4-8:

- The system tries to allocate the wanted values.
- If the managed system cannot provide the minimum amount of memory, the partition does not start.
- You cannot dynamically increase the amount of memory in a partition to more than the defined maximum. If you want more memory than the maximum, the partition must be stopped and reactivated with an updated profile (not only restarted).

6. Storage adapter considerations.

   A decision is needed for the type of storage to assign to the VIOS partition for installation. If the decision is to use internal disks, then two types of adapters can be assigned based on the specification of the Power server: SAS disks (attached to a SAS adapter) or NVMe disks (attached to an NVMe adapter).

   For internal storage, select the corresponding adapter from the list of adapters on the Physical I/O Adapter window, as shown in Figure 4-11 on page 148.

If the VIOS starts from a storage area network (SAN) LUN, more preparation is needed before the physical I/O adapter assignment is done:

- Locate the physical Fibre Channel (FC) adapter and the worldwide port name (WWPN) by using the HMC Enhanced GUI.
- The SAN switch requires zoning the physical FC adapter port WWPN with the storage target port WWPN or as recommended by the vendor's documentation.
- The SAN storage requires LUN masking for the physical FC port WWPN.

The following instructions describe how to locate the physical FC port WWPN by using the HMC Enhanced GUI:

a. Select **Resources**.

b. Select **All Systems**.

c. Click the managed system name.

d. Select **Processor, Memory, I/O** from th left pane.

e. Expand **Physical I/O Adapters**.

f. Expand the adapter that is selected to be assigned to the VIOS LPAR, which shows the adapter ports.

g. Click the information icon to locate the port's WWPN (see Figure 4-9).

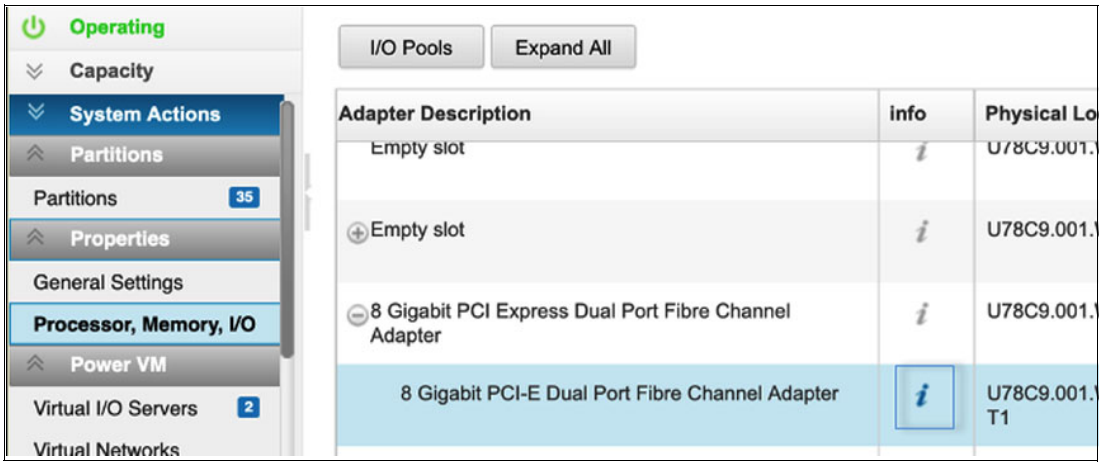


Figure 4-9 Processor, Memory, and I/O on the HMC Enhanced GUI showing a physical FC adapter with dual ports

The window that opens shows the WWPN information, as shown in Figure 4-10 on page 147.

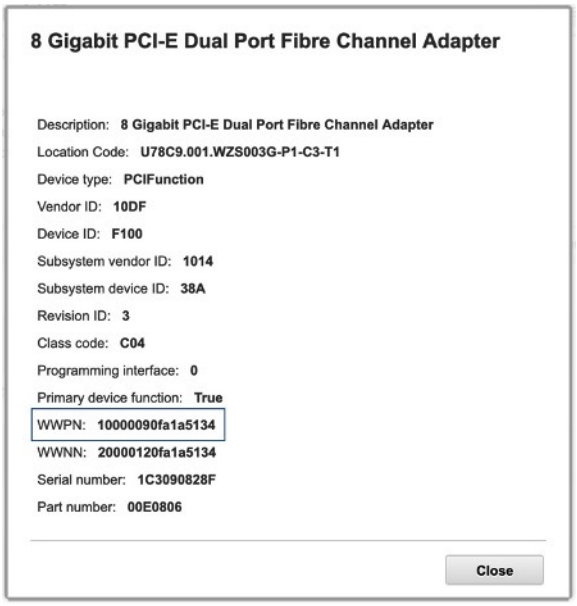


Figure 4-10 FC port informational Vital Product Data showing the physical WWPN on the port

Alternatively, you can the HMC CLI to find the same information. Use an Secure Shell (SSH) to connect to the HMC IP address or the fully qualified domain name (FQDN) by using a `hmcsuperuser` (like `hscroot`) username and password. Run the following command:

```
~> lshwres -r io -m <Managed system name> --rsubtype slotchildren -F
phys_loc,description,wwpn
```

Here is example output of the command:

```
U78C9.001.WZS003G-P1-C3-T1,8 Gigabit PCI-E Dual Port Fibre Channel
Adapter,10000090fa1a5134
```

Note: The managed system name can be listed by using the following HMC command:

```
~> lssyscfg -r sys -F name
```

The physical FC adapter port WWPN can be used to continue the zoning from the SAN switch and mapping or masking from the SAN storage.

Notes:

- If the WWPNs are not shown on the HMC Enhanced GUI or the HMC CLI, then a hardware discovery is required. For more information, see lshwres command returns “No results were found” when displaying WWPN information of physical adapters on HMC, found at:

  https://www.ibm.com/support/pages/lshwres-command-returns-%E2%80%9Cno-results-were-found%E2%80%9D-when-displaying-wwpn-information-physical-adapters-hmc

- SAN switch zoning and SAN storage LUN masking are beyond the scope of this book. For more information about zoning and mapping, see your SAN product vendor’s documentation.

7. Select the physical I/O adapter to assign to the VIOS LPAR.

   To successfully create an LPAR, assign a storage adapter and network adapter. For more information, see Minimum hardware configuration requirements for logical partitions, found at:

   https://www.ibm.com/docs/en/power10?topic=partitions-minimum-hardware-configuration-requirements-logical

Note: For VIOS installation by using a Universal Serial Bus (USB) CD or DVD drive or USB flash, you must assign the USB physical adapter to the VIOS with a description like “Universal Serial BUS UHC”, as shown in Figure 4-11.

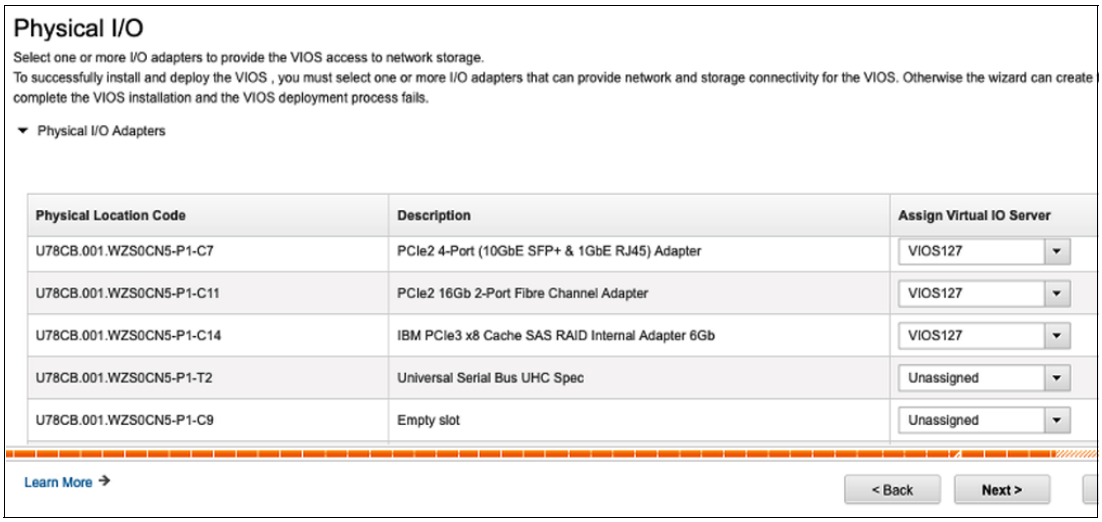


Figure 4-11 Physical I/O adapter assignment on the HMC Enhanced GUI during VIOS LPAR creation

Click **Next**.

8. Review the configuration summary, as shown in Figure 4-12.

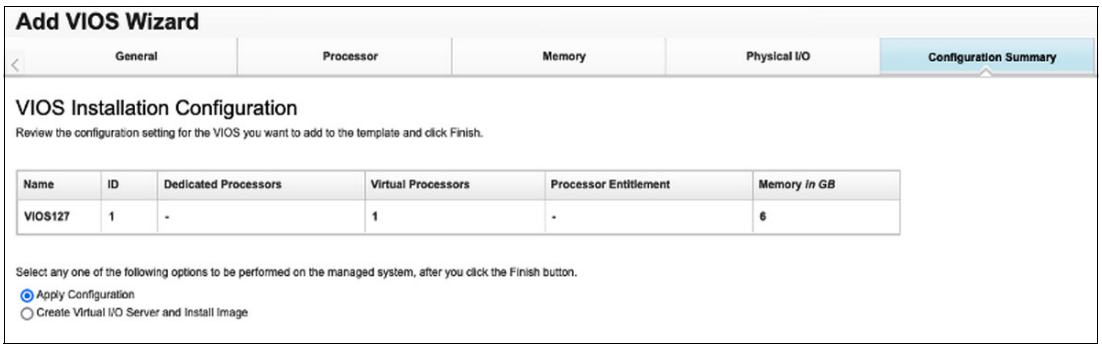


Figure 4-12 Summary of configurations that are selected on the HMC Enhanced GUI

Note: After the LPAR is created, you must enable **Save configuration changes to profile**. For more information, see Saving Configuration Changes To Profile, found at:

https://www.ibm.com/support/pages/saving-configuration-changes-profile

4.2.3 VIOS installation methods

On the Configuration summary page in Figure 4-12 on page 148, there are two options: **Apply Configuration** and **Create Virtual I/O server and install Image**.

The **Apply Configuration** option creates the VIOS partition without VIOS installation. In this case, installation from console is required.

The **Create Virtual I/O Server and Install Image** option creates the partition and provides multiple options for VIOS installation. The most common options are to install the VIOS from the Network Installation Manager (NIM) or from the HMC repository.

The next section shows the implementations for these two options.

VIOS installation media

As a best practice, download the VIOS installation media from IBM Entitled Systems Support (IBM ESS). Before the VIOS installation media is downloaded from IBM ESS, use the Fix Level Recommendation Tool (FLRT) to identify the latest and recommended VIOS version to use. For more information about the FLRT tool, see PowerVM Virtual I/O Server on FLRT Lite, found at:

https://esupport.ibm.com/customercare/flrt/liteTable?prodKey=vios

For more information about how to obtain the VIOS installation media from IBM ESS, see How to Obtain Installation Software for PowerVM Virtual I/O Server, found at:

https://www.ibm.com/support/pages/how-obtain-installation-software-powervm-virtual-io-server

VIOS installation during LPAR creation

In Figure 4-12 on page 148, select **Create Virtual I/O Server and Install Image**, and then click **Finish**.

The VIOS LPAR can be created and installed from the Add VIOS wizard by selecting **Create Virtual I/O Server and install Image**. The wizard creates the VIOS LPAR and provides four installation options. The most common installation options are as follows:

- NIM server
- Management Console Image

NIM server

If you have an existing environment and a NIM server is configured, then the NIM IP address can be entered to start the installation of VIOS from NIM. For more information about how to set up NIM for an existing Power server environment, see NIM Setup Guide, found at:

https://www.ibm.com/support/pages/nim-setup-guide

Figure 4-13 shows the configuration window for the NIM server installation option. Here, you can enter the NIM server IP address and the IP address information for the physical network adapter port that is cabled, which can communicate with the NIM server.

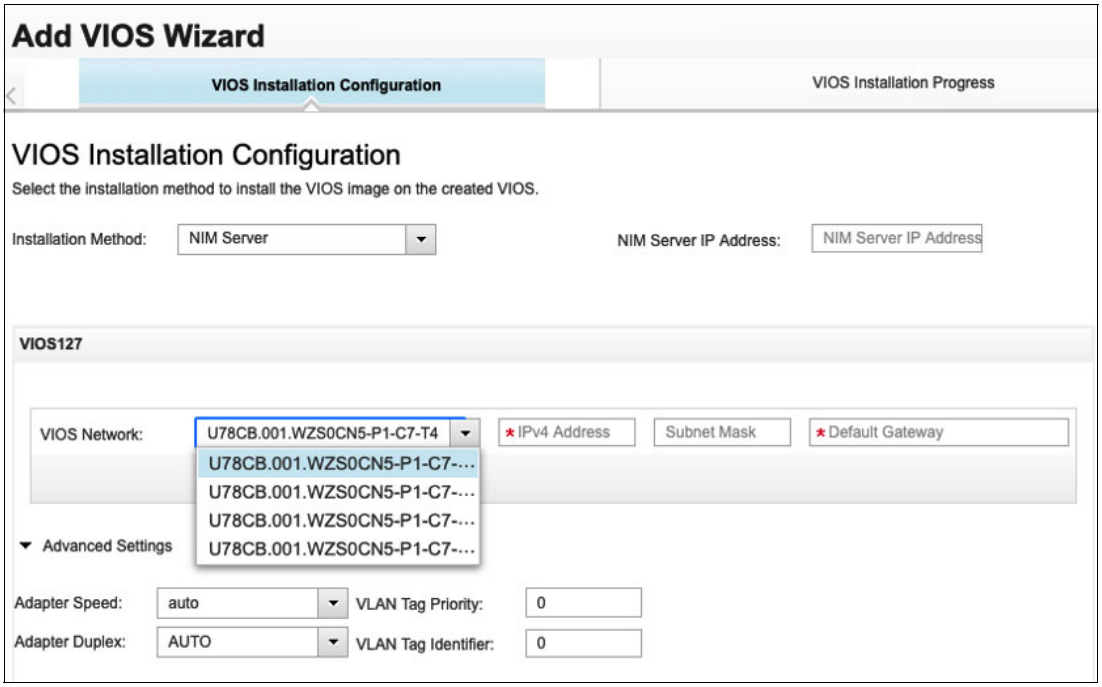


Figure 4-13 NIM Server configuration: Enter the NIM IP address to start the installation from the NIM server

Select **Next** → **Start** → **Accept License** (if requested) → **Finish**.

Management Console Image

Use this option to install the VIOS from an image that is stored in the HMC repository. Figure 4-14 on page 151 shows the configuration page and the VIOS image that is available in the HMC repository. Select the Management Console IP address, the VIOS image that is stored in the HMC repository, and the IP configuration on the VIOS network port. Select **Next** → **Start** → **Accept License** (if requested) → **Finish**.

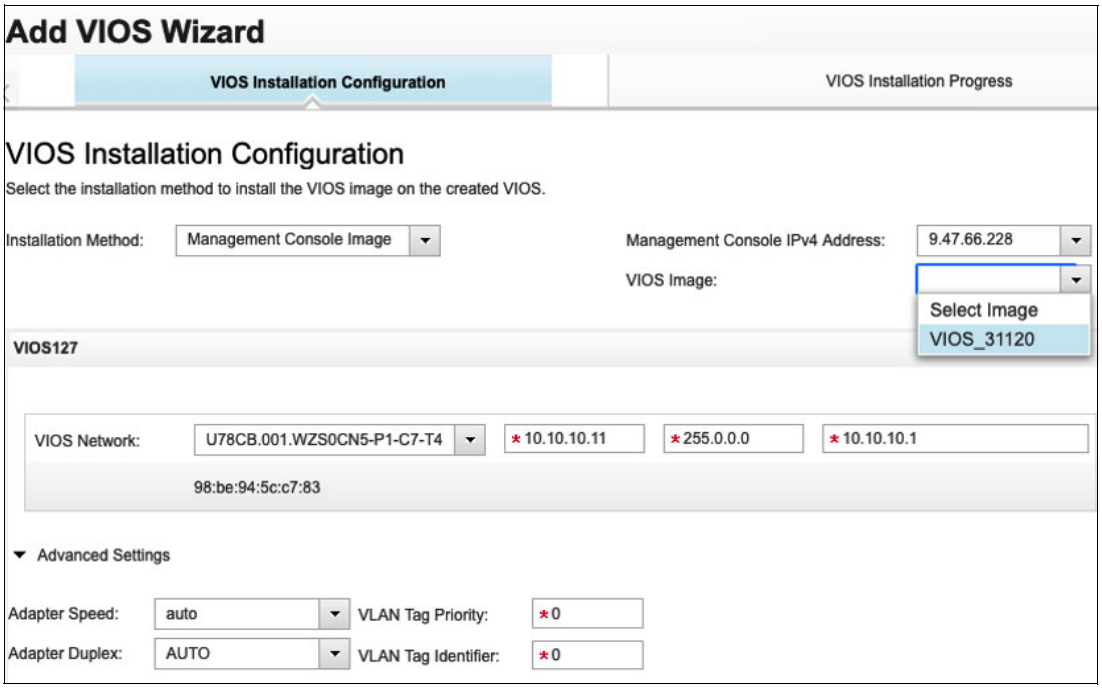


Figure 4-14 Management Console Image installation: VIOS 3.1.1.20 stored in the HMC repository

For more information about how to add a VIOS installation image to the HMC repository, see Manage Virtual I/O Server Image Repository, found at:

https://www.ibm.com/docs/en/power10/7063-CR2?topic=images-manage-virtual-io-server-image-repository

VIOS installation from a console

Select **Apply Configuration**, as shown in Figure 4-12 on page 148, and click **Finish**. It creates a VIOS LPAR with the assigned resources.

The VIOS LPAR can be installed from the console by activating the VIOS LPA, as shown in Figure 4-15.

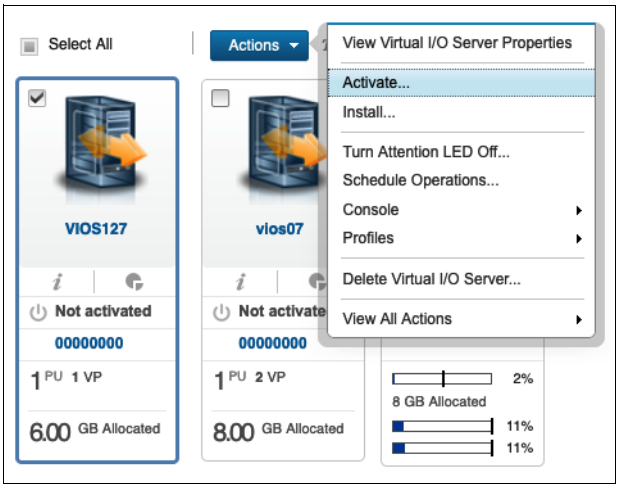


Figure 4-15 Activating the VIOS LPAR

Select **System Management Services** for Boot mode, as shown in Figure 4-16, and click **Finish**.

Note: You can select **`open vterm`**, which opens a Java virtual terminal application to manage the VIOS System Management Services (SMS) configuration. However, a better option is to SSH into the HMC by using your preferred SSH client terminal application and run the command **`vtmenu`**. Type the managed system number and press Enter. Then, type the VIOS LPAR number and press Enter. Now, you have a console terminal to the VIOS LPAR.

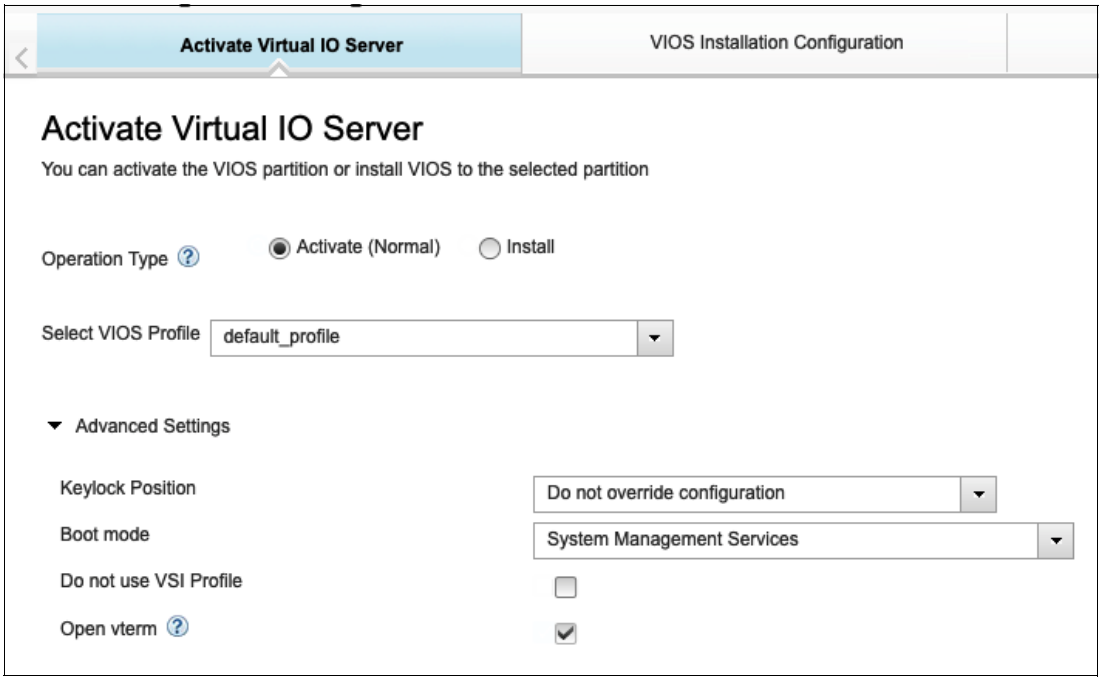


Figure 4-16 System Management Services selected as the boot mode

VIOS can be installed by using a USB flash drive, USB CD/DVD drive, or NIM:

- For USB flash drive preparation and th installation of VIOS, see PowerVM VIOS 3.1 Installation Using USB Flash Drive, found at:

  https://www.ibm.com/support/pages/powervm-vios-31-installation-using-usb-flash-drive

- For a USB CD or DVD drive installation of VIOS from SMS, follow these steps:

  a. Make sure that the disc is inserted in the USB CD or DVD drive.

  b. Select `5. Select Boot Options` and then press Enter.

  c. Select `1. Select Install/Boot Device` and then press Enter.

  d. Select `7. List all Devices`, look for the CD-ROM (press `n` to get to the next page if required), enter the number for the CD-ROM, and then press Enter.

  e. Select `2. Normal Mode Boot` and then press Enter.

  f. Confirm your choice by selecting `1. Yes` and then pressing Enter. ?

  g. You are prompted to accept the terminal as `console` and then to select the installation language.

h. You are presented with the installation menu. It is best to check the settings (option 2) before proceeding with the installation. Check whether the selected installation disk is correct.

i. When the installation procedure finishes, use the **padmin** username to log in. On initial login, you are prompted to specify a password. There is no default password.

j. After a successful login, you are under the VIOS CLI.

k. Enter `a` (and press Enter) to accept the Software Maintenance Agreement terms, then run the following command to accept the license:

```
$ license -accept
```

The VIOS installation is complete.

- For NIM preparation and VIOS installation, complete the following steps:

  a. Download the VIOS installation media (ISO files) from the IBM ESS website.

  b. Extract the `mksysb` from the ISO files.

  c. Define `mksysb` and SPOT.

  d. Allocate resources for network boot (`nim_bosinst`).

  For more information, see How to set up NIM for VIOS installation, found at:

  https://www.ibm.com/support/pages/node/6829921

4.2.4 VIOS initial configuration

This section describes the basic configuration that is required after the VIOS installation, which includes the following items:

- VIOS rules
- Adding an IP configuration
- Configuring Network Time Protocol
- Configuring name resolution

VIOS rules

The device settings on VIOS play a critical role in optimizing the performance of your client partitions. Tune and modify the device settings on VIOS according to the recommended settings. You must have consistent settings across multiple VIOSs.

VIOS rules management provides an effective way to tune device settings, and it also provides a simple way to replicate the changes across multiple VIOSs.

VIOS rules management provides predefined default device settings that are based on the best practices values for VIOS. These device settings on VIOS can be managed and customized as required by using VIOS rules management.

VIOS rules management provides the flexibility to collect, verify, and apply device settings.

VIOS rules management consists of two rules files. These rules files are created in XML format.

- Default rules file

  This file contains the critical suggested device rules that follow VIOS best practices. This file is included with VIOS with read-only permissions. Figure 4-17 shows the sample default rules file.

- Current rules file

  This file contains the current VIOS system settings based on the default rules. The user can use this file to customize device settings. This file can be modified by using the **rules** command.

```
<?xml version="1.0" encoding="UTF-8"?>
<!-- When VIOS level changes, the value of ioslevel needs to change manually -->
<Profile origin="get" version="3.0.0" date="2012-10-05T00:00:00Z">
 <Catalog id="devParam.disk.fcp.mpioosdisk" version="3.0">
  <Parameter name="reserve_policy" value="no_reserve" applyType="nextboot" reboot="true">
   <Target class="device" instance="disk/fcp/mpioosdisk"/>
  </Parameter>
 </Catalog>
 <Catalog id="devParam.disk.fcp.mpioapdisk" version="3.0">
  <Parameter name="reserve_policy" value="no_reserve" applyType="nextboot" reboot="true">
   <Target class="device" instance="disk/fcp/mpioapdisk"/>
  </Parameter>
 </Catalog>
 <Catalog id="devParam.disk.fcp.nonmpiodisk" version="3.0">
  <Parameter name="reserve_policy" value="no_reserve" applyType="nextboot" reboot="true">
   <Target class="device" instance="disk/fcp/nonmpiodisk"/>
  </Parameter>
 </Catalog>
```

Figure 4-17 Sample default rules file

VIOS rules can be deployed, verified, and captured by using the **rules** command.

For more information, see Managing VIOS rules files, found at:

https://www.ibm.com/docs/en/power10/9080-HEX?topic=management-managing-vios-rules-files

The **rulescfgset** command is an interactive tool to guide a user that is deploying current rules at a user's direction. The command helps to simplify the rules deployment management process. Figure 4-18 shows a sample run of this command.

```
padmin@vios05:/home/padmin>rulescfgset
The most recent software update introduces the concept of Rules. Rules can be
created, modified, listed, deleted and deployed for specific system settings.
See IBM documentation for more details on rules.  The software update includes
the best practice settings from IBM as factory default rules. Do you want to
deploy default rules on top of the current system settings now [y/N]? y

bosboot: Boot image is 61468 512 byte blocks.
The new device settings have been deployed successfully
```

Figure 4-18 Sample run of the rulescfgset command

For more information, see the **rulescfgset** command, found at:

https://www.ibm.com/docs/en/power10/9080-HEX?topic=commands-rulescfgset-command

Adding an IP configuration

If the VIOS is installed from the console, then an IP configuration is required on the network adapter. Network communication is required for a Resource Monitoring and Control (RMC) connection with the HMC. To configure the network adapter interface on the VIOS, complete the following steps:

1. List the Ethernet devices that are configured on the VIOS:

```
$ lsdev -type adapter
ent0 Available   4-Port Gigabit Ethernet PCI-Express Adapter (e414571614102004)
ent1 Available   4-Port Gigabit Ethernet PCI-Express Adapter (e414571614102004)
ent2 Available   4-Port Gigabit Ethernet PCI-Express Adapter (e414571614102004)
ent3 Available   4-Port Gigabit Ethernet PCI-Express Adapter (e414571614102004)
```

   `ent0` is the physical Ethernet adapter port that is cabled. This example uses the network port `ent0` to configure an IP on the interface of this port (`en0`).

2. Configure the IP address on the port interface `en0` by running the following command:

```
$ mktcpip -hostname <Desired VIOS hostname> -interface en0 -inetaddr <IP
Address> -netmask <network mask> -gateway <gateway>
```

   Alternatively, you can run **`cfgassist`** from the VIOS CLI, which is a curses-based text interface to perform system management functions, as shown in Figure 4-19.

```
Move cursor to desired item and press Enter.

  Set Date and TimeZone
  Change Passwords
  Set System Security
  VIOS TCP/IP Configuration
  Install and Update Software
  Storage Management
  Devices
  Performance
  Role Based Access Control (RBAC)
  Shared Storage Pools
  Electronic Service Agent
```

Figure 4-19 The cfgassist menu

3. Select `en0` to configure the IP address on `en0`, as shown Figure 4-20.

```
                        Available Network Interfaces

 Move cursor to desired item and press Enter.

   en0            Standard Ethernet Network Interface
   en11           Standard Ethernet Network Interface
   en4    04-00   Standard Ethernet Network Interface
   en5    04-01   Standard Ethernet Network Interface
   en6    04-02   Standard Ethernet Network Interface
   en7    04-03   Standard Ethernet Network Interface
   en8            Standard Ethernet Network Interface
   en9            Standard Ethernet Network Interface
   et0            IEEE 802.3 Ethernet Network Interface
   et11           IEEE 802.3 Ethernet Network Interface
   et4    04-00   IEEE 802.3 Ethernet Network Interface
   et5    04-01   IEEE 802.3 Ethernet Network Interface
   et6    04-02   IEEE 802.3 Ethernet Network Interface
   et7    04-03   IEEE 802.3 Ethernet Network Interface
   et8            IEEE 802.3 Ethernet Network Interface
   et9            IEEE 802.3 Ethernet Network Interface

 F1=Help                F2=Refresh             F3=Cancel
 F8=Image               F10=Exit               Enter=Do
 /=Find                 n=Find Next
```

Figure 4-20 Available network interfaces

4. Type or select values in the entry field, as shown in Figure 4-21.

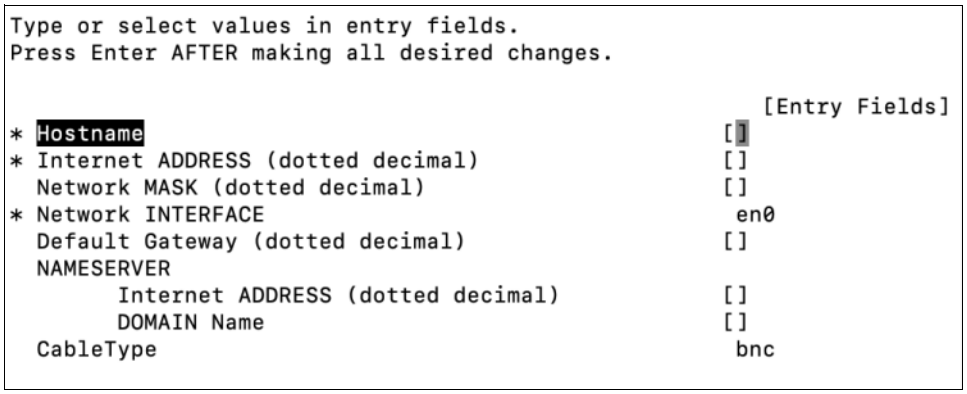


Figure 4-21 IP configuration from cfgassist

Configuring Network Time Protocol

Synchronized timing is important for error logging and various monitoring tools. As a best practice, configure the VIOS as a Network Time Protocol (NTP) client by following the steps in Configuring NTP in PowerVM VIOS, found at:

https://www.ibm.com/support/pages/configuring-ntp-powervm-vios

Configuring name resolution

You must configure name resolution on VIOS because anytime a query is run from the HMC, a call is made to all VIOS on the managed system to get configuration details. If the name resolution is not correctly configured, the HMC query might fail.

When the HMC sends a query to the VIOS, it might attempt to resolve hostnames and IP addresses by using the following sources:

- BIND/DNS (domain name server)
- The local `/etc/hosts` file

By default, it first attempts resolution by using BIND/DNS. If the `/etc/resolv.conf` file does not exist or if BIND/DNS cannot find the entry, then the local `/etc/hosts` file is searched.

The default order can be overridden by creating the configuration file `/etc/netsvc.conf` and specifying the order. For local name resolution (recommended), the following entry can be appended to the `/etc/netsvc.conf` file:

```
hosts=local,bind4
```

The `/etc/hosts` file must include the VIOS IP, FQDN, and short name. Append the following entry in the `/etc/hosts` file with the following format:

```
<VIOS IP Address> <Fully Qualified Domain Name(FQDN)> <alias>
```

Note:

- To list the VIOS IP address that is configured on network interfaces, run the following command:

  ```
  $ netstat -state
  ```

- To view the current VIOS hostname, run the following command:

  ```
  $ lsdev -dev inet0 -attr |grep hostname
  ```

- To change the hostname, run the following command (a restart is not required):

  ```
  $ chdev -dev inet0 -attr hostname=<preferred VIOS hostname>
  ```

4.3 Network configuration

This section describes network virtualization configuration at the managed system level on the HMC GUI.

Two types of networking implementation are described in this section:

- Virtual network
- Single-root I/O virtualization (SR-IOV)

Network virtualization concepts are described in 2.4, “Network virtualization” on page 45. SR-IOV is described in 2.4.3, “Single-root I/O virtualization” on page 47.

For network virtualization planning considerations, see 3.6, “Network virtualization planning” on page 118. For SR-IOV planning, see 3.6.5, “SR-IOV planning” on page 134.

4.3.1 Virtual network configuration

Virtual networks can be created on managed systems by using HMC. Virtual networks are accessed by using Virtual Ethernet Adapters (VEAs). One or more VEAs can be configured on the VIOS and then bridged to physical Ethernet adapters by forming a Shared Ethernet Adapter (SEA) on VIOS.

As a best practice, use a dual-VIOS LPAR in the environment to leverage the high availability (HA) feature that the virtual network provides between the dual VIOSs.

The following resources describe how to create a virtual network from the HMC Enhanced GUI:

- IBM VIOS: Create Virtual Network with VLAN tagging - Load sharing configuration, found at:

  https://www.ibm.com/support/pages/node/1097038

- Managing virtual networks, found at:

  https://www.ibm.com/docs/en/power10?topic=systems-managing-virtual-networks

4.3.2 Single-root I/O virtualization configuration

SRI-IOV shared mode can be enabled from the HMC Enhanced GUI. The SR-IOV-capable physical Ethernet adapter is assigned to the PowerVM hypervisor (PHYP). The adapter can be shared between multiple LPARs concurrently (the virtual function (VF) can be assigned to client LPARs and to the VIOS).

To check the readiness and configure the capacity percentage and VF, complete the following steps:

1. Check the Licensed Capabilities of the managed system and ensure that they are SR-IOV-capable, as shown on Figure 4-22.

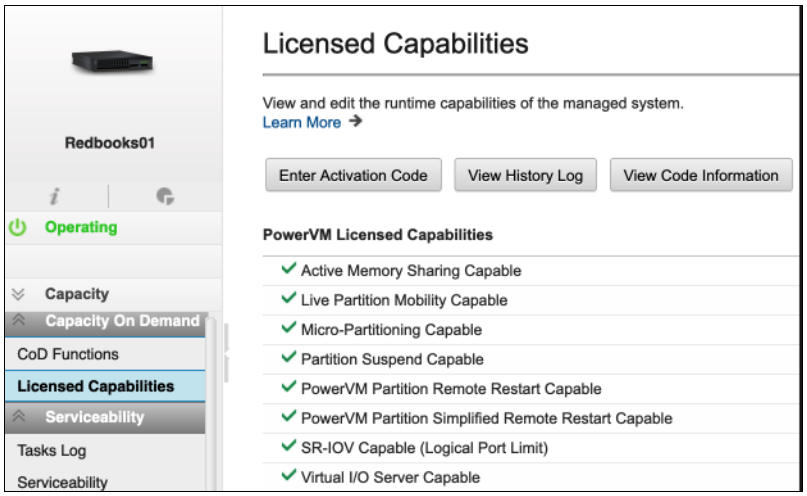


Figure 4-22 SRI-OV-capable on Licensed Capabilities

2. If the SR-IOV Capable feature is enabled, then check whether the physical network adapter is SR-IOV-capable. From the HMC, select the managed system name and then select **Processor, Memory, I/O**. Expand **Physical I/O Adapters**, and check whether the physical network adapter is SR-IOV-capable, as shown in Figure 4-23. This figure shows that the adapter is not SR-IOV-capable. You must use an SR-IOV-capable adapter that is an unassigned adapter.

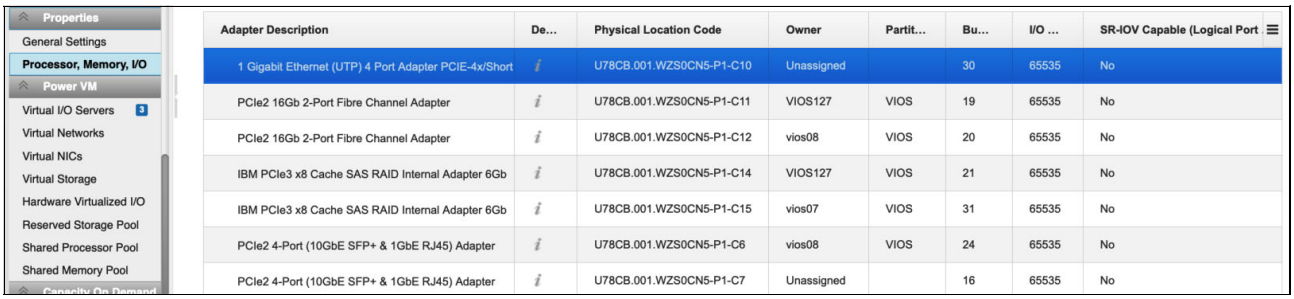


Figure 4-23 Finding an SR-IOV-capable adapter

3. Modify the SRI-OV adapter and change the mode from dedicated to shared by following the steps in Modifying SR-IOV adapters, found at:

   https://www.ibm.com/docs/en/power10/9786-22H?topic=adapters-modifying-sr-iov

4. Modify the SR-IOV physical ports by following the steps in Modifying SR-IOV physical port settings, found at:

   https://www.ibm.com/docs/en/power10/9786-22H?topic=adapters-modifying-sr-iov-physical-port-settings

5. Add logical ports (LPs) to the partition (either VIOS or client LPARs) by following the steps in Adding SR-IOV logical ports, found at:

   https://www.ibm.com/docs/en/power10?topic=settings-adding-sr-iov-logical-ports

4.4 Creating and installing a client LPAR

This section describes the creation of a client LPAR and assigning resources to it.

4.4.1 Creating a client LPAR

To create a client LPAR, click **All Systems**, click the management system name, click **Partitions** from the left pane, and click **Create partition**, as shown in Figure 4-24.

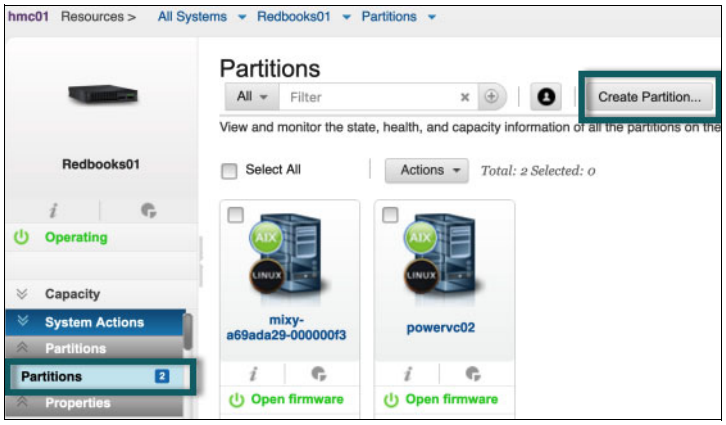


Figure 4-24 Client LPAR creation

Enter the Basic Partition Configuration details, processor, and memory configuration for the client LPAR, as shown in Figure 4-25. Then, click **OK** to create a client LPAR.

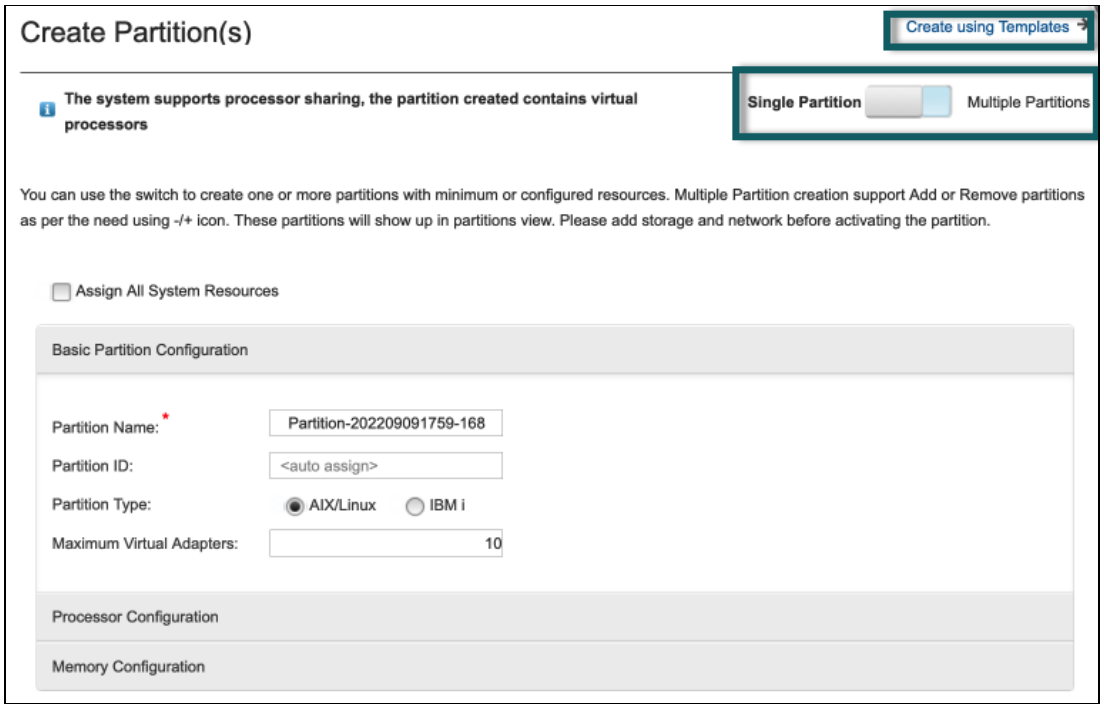


Figure 4-25 Client LPAR Create Partitions

Note: Each virtual adapter slot consumes a small amount of memory. Therefore, as a best practice, assign a reasonable number of Maximum Virtual Adapters that is based on the requirement for client storage and network adapters.

After the LPAR is created, it is important to enable **Save configuration changes to profile**. For more information, see Saving Configuration Changes To Profile, found at:

`https://www.ibm.com/support/pages/saving-configuration-changes-profile`

Figure 4-25 shows more options as follows:

- **Single Partition**: Select this option to create a single partition.
- **Multiple partitions**: Use this option to create several partitions after you enter the processor and memory configuration. This option creates partitions without assigning any network or storage adapters, which can be assigned later to the partitions.
- **Create from template**: This option allows the creation of client LPARs from a predefined template. This template contains the configuration and resources, for example, processor, memory, virtual network, and virtual storage, and other features that can be set up based on your preferences. To create a partition from a template, see Creating a logical partition by using a template, found at:

  `https://www.ibm.com/docs/en/power10?topic=templates-creating-logical-partition-by-using-template`

To customize your own template, you can copy an existing predefined template, go to the HMC Enhanced GUI. Select **HMC management** → **Templates and OS Images**, as shown in Figure 4-26. Select **Partition**, and select one of the predefined templates. Click **Copy** and customize the template based on your preferences.

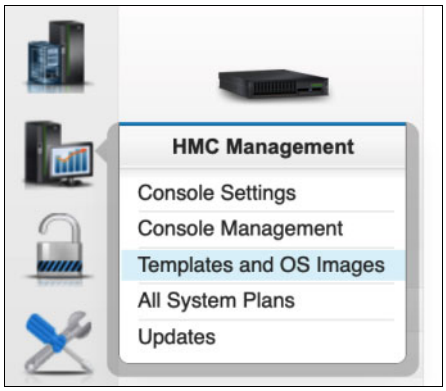


Figure 4-26 Templates and OS Images

4.4.2 Capturing and deploying VMs with PowerVC

In a PowerVC managed environment, VMs can be provisioned and installed by using the capture and deploy feature. For more information, see the following resources:

- Capturing a virtual machine, found at:

  https://www.ibm.com/docs/en/powervc/2.1.0?topic=images-capturing-virtual-machine

- Deploying captured or imported images, found at:

  https://www.ibm.com/docs/en/powervc/2.0.3?topic=images-deploying-captured-imported

4.4.3 Client LPAR storage configuration

Two types of storage adapters can be assigned to a client LPAR:

- Virtual SCSI (vSCSI)

  The disk on the client is backed by a logical or a physical device on the VIOS, which can be a physical volume, a logical volume, or logical unit (LU).

  vSCSI is described in 2.3.1, “Virtual SCSI” on page 42. For planning considerations, see 3.5.1, “Virtual SCSI planning” on page 101.

  For more information, see Virtual SCSI, found at:

  https://www.ibm.com/docs/en/power10?topic=overview-virtual-scsi

- N_Port ID Virtualization (NPIV)

  NPIV allows multiple LPARs to access SAN storage through the same physical FC adapter. For NPIV, SAN LUNs are presented directly to the client LPAR. VIOS is a pass-through between the SAN storage and the client operating system, so no LUNs are assigned to the VIOS.

  NPIV is described in 2.3.2, “Virtual Fibre Channel” on page 43. For planning considerations, see 3.5.2, “Virtual Fibre Channel planning” on page 104.

For more information, see Virtual Fibre Channel, found at:

https://www.ibm.com/docs/en/power10?topic=overview-virtual-fibre-channel

To assign a storage adapter to a client LPAR, follow these steps:

1. Click **All Systems**.
2. Click the managed system name.
3. Click **Partitions**.
4. Click the name of the newly created client LPAR.
5. Click **Virtual Storage**, as shown on Figure 4-27.

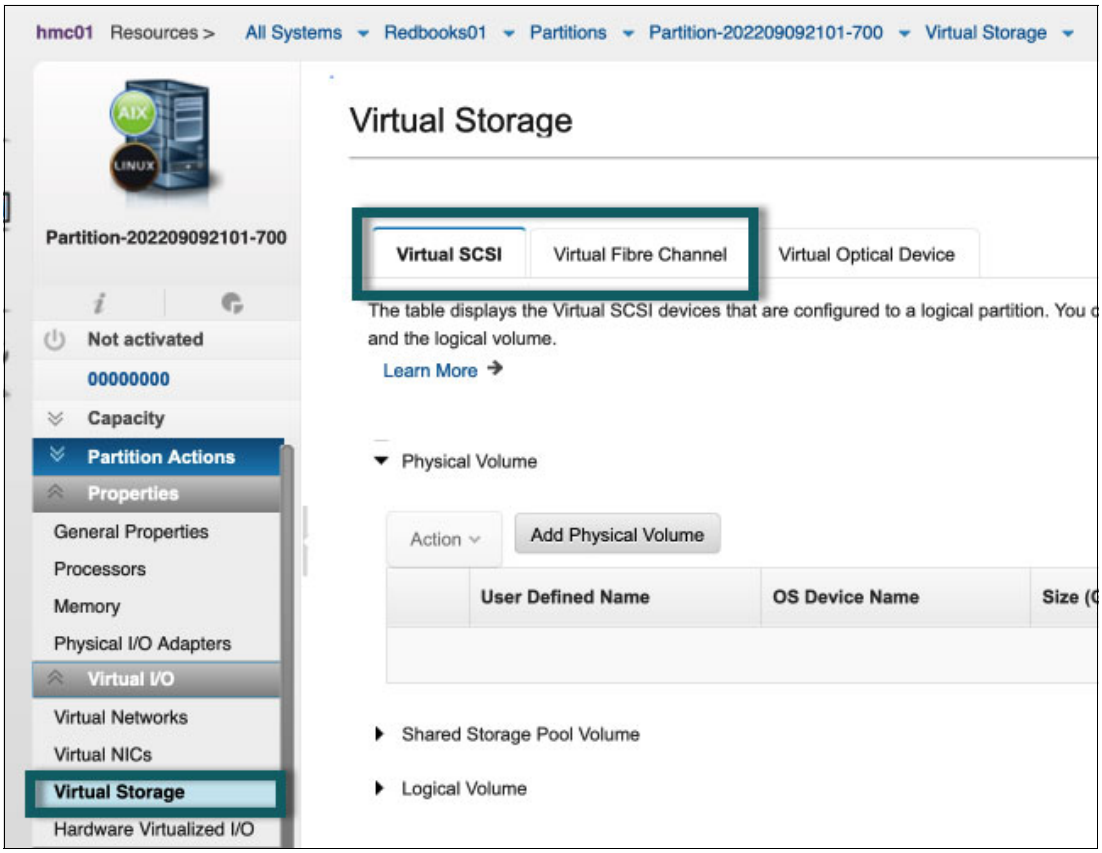


Figure 4-27 Client virtual storage configuration

For more information about adding vSCSI-backed storage (physical volume, logical volume and shared storage pool (SSP) LUs), see Managing virtual SCSI resources for a partition, found at:

https://www.ibm.com/docs/en/power10/9043-MRX?topic=view-managing-virtual-scsi-resources-partition

For more information about adding virtual Fibre Channel (NPIV), see How to DLPAR Virtual Fibre Channel Adapters Using the HMC Enhanced GUI, found at:

https://www.ibm.com/support/pages/how-dlpar-virtual-fibre-channel-adapters-using-hmc-enhanced-gui

Notes:

- For NPIV, after the virtual Fibre Channel (VFC) adapter is created from the HMC Enhanced GUI, two virtual WWPNs are generated for each adapter. The primary WWPN must be zoned on the SAN switch and LUNs must be masked on the SAN storage.
- If LPM is used between Power servers (managed systems), you must zone the secondary WWPN on the SAN switch the same way that the primary WWPN is zoned. The same LUNs also must be masked to the secondary WWPN.
- Usually, SAN administrators prefer to scan the SAN switch to discover the client WWPN in preparation to zone it. If the client WWPNs cannot be discovered, then log in from the HMC Enhanced GUI. From the client LPAR, complete these steps:
  a. Select **Virtual Storage**.
  b. Select the **Virtual Fibre Channel** tab.
  c. Click **Log In.**
  d. After zoning is completed, click **Log Off**.

4.4.4 Client LPAR network configuration

Several types of client networking can be added based on the user's preference:

- Virtual network
- SRI-OV
- Virtual Network Interface Controllers (vNICs)
- Hybrid Network Virtualization (HNV)

Virtual network

The following steps describe how to connect a client LPAR to a virtual local area network (VLAN) when a virtual network is used. This procedure creates a VEA on the client LPAR.

1. Click **All Systems**.
2. Click the managed system name.
3. Select **Partitions**.
4. Click the client LPAR name.
5. Click **Virtual Networks** on the left pane.

6. Click **Attach Virtual Network** and select the preferred VLAN to use on the client LPAR, as shown in Figure 4-28.
7. Click **OK**.

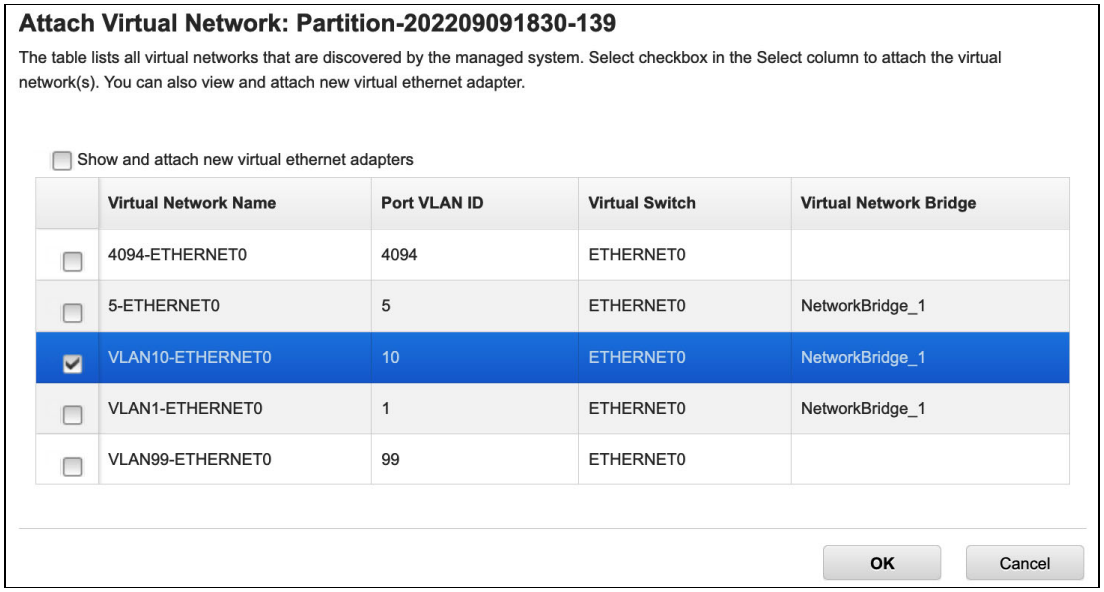


Figure 4-28 Attach Virtual Network window

SRI-OV

If SRI-OV logical port (LP) is selected as the network type for the client LPAR, see Adding SR-IOV logical ports, found at:

https://www.ibm.com/docs/en/power10?topic=settings-adding-sr-iov-logical-ports

vNIC

If vNIC is selected as the network type for the client LPAR, for more information about how to create vNIC for the client LPAR, see vNIC Functionality Guide, found at:

https://www.ibm.com/support/pages/vnic-functionality-guide

Hybrid Network Virtualization

If the client LPAR participates in LPM, and you prefer to use SRI-OV for client LPAR networking, then you can use HNV. For more information, see Hybrid Network Virtualization - Using SR-IOV for Optimal Performance and Mobility, found at:

https://community.ibm.com/community/user/power/blogs/charles-graham1/2020/06/19/hybrid-network-virtualization-using-sr-iov-for-opt

4.4.5 Installing the client operating system

This section describes the common installation methods to install an operating system on a client LPAR. These methods can be categorized as follows:

- Using a virtual optical device: A vSCSI adapter that is backed by a logical or physical device on VIOS. It can be a USB flash drive, USB optical drive (DVD-RAM or DVD-ROM), or a media repository ISO file.
- Using a physical device: USB flash drive or USB optical drive. The physical USB adapter must be assigned to the client LPAR. This method is *not recommended* if you plan to migrate the client to another managed system.
- Network Installation.

Virtual optical device

With the virtual optical device installation, the vSCSI adapter mapping is created between the VIOS and the client LPAR.

A virtual target device (VTD) is created under the vhost on VIOS. You can map either a physical USB optical device, a USB flash drive, or an image from the repository to the vhost.

For more information about the procedure to create the virtual optical device from the HMC Enhanced GUI (Figure 4-29), see Adding virtual optical devices, found at:

https://www.ibm.com/docs/en/power10?topic=assignment-adding-virtual-optical-devices

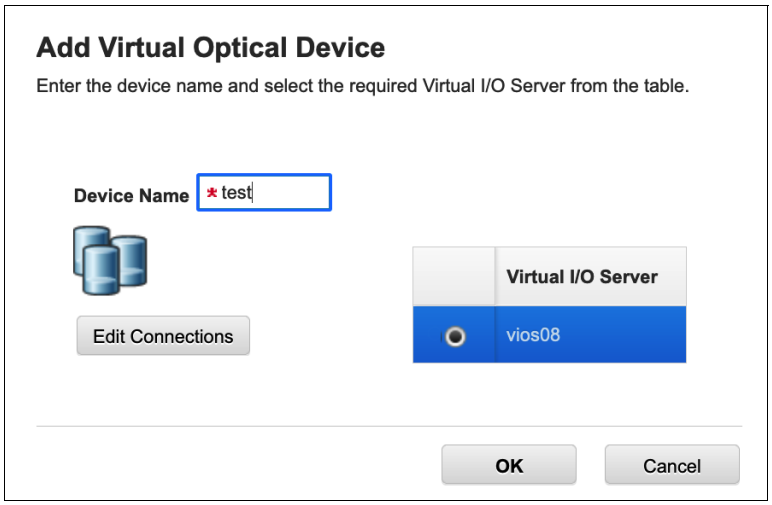


Figure 4-29 Add Virtual Optical Device

SSH in to VIOS and run the **lsmap -all** command to check whether the vhost and the test VTD were created, as shown in Figure 4-30.

```
padmin@vios08:/home/padmin>lsmap -all
SVSA            Physloc                                      Client Partition ID
--------------- -------------------------------------------- ------------------
vhost2          U8247.21L.2142B2A-V2-C10                     0x00000006

VTD                   test
Status                Available
LUN                   0x8100000000000000
Backing device
Physloc
Mirrored              N/A

padmin@vios08:/home/padmin>
```

Figure 4-30 The lsmap command output

- If a USB optical device (DVD-RAM or DVD-ROM) exists on VIOS, you can view the optical device name by running the following command:

  ```
  $ lsdev -type optical
  ```

  For more information, see Moving the DVD-RAM Between LPARs using the VIO server, found at:

  https://www.ibm.com/support/pages/moving-dvd-ram-between-lpars-using-vio-server

- If the USB flash drive is assigned to the VIOS, you can view the USB flash drive by running the following command:

  ```
  $ lsdev |grep -i usb
  ```

  For more information, see Using and taking advantage from USB devices and AIX, found at:

  https://www.ibm.com/support/pages/using-and-taking-advantage-usb-devices-and-aix

- If the installation is being made from a media repository, you can view the images in the image repository by running the **lsrep** command on VIOS, as shown in Figure 4-31.

```
padmin@vios08:/home/padmin>lsrep
Size(mb) Free(mb) Parent Pool         Parent Size      Parent Free
   20397    17830 rootvg                   61376            16576

Name                                                  File Size Optical         Access
RHEL-8.2.0-20200404.0-ppc64le-dvd1                         2566 None            rw
```

Figure 4-31 The lsrep command to list the available repository images that can be assigned to vhost2

For more information, see How to configure a VIOS Media Repository/Virtual Media Library, found at:

https://www.ibm.com/support/pages/how-configure-vios-media-repositoryvirtual-media-library-ex-aix-installrestore

Alternatively, for HMC GUI instructions about managing a media repository, see the following resources:

- Adding or removing a media library, found at:

  https://www.ibm.com/docs/en/power10?topic=libraries-adding-removing-media-library

- Adding or removing media files from a media library, found at:

  https://www.ibm.com/docs/en/power10/9105-22A?topic=libraries-adding-removing-media-files-from-media-library

Regardless of the selected virtual device method, the command to map a device or file to vhost2 remains the same as follows:

```
$ mkvdev -vdev <device name> -vadapter <vhost#>
```

For example:

```
$ mkvdev -vdev <cd0| usbms0 | RHEL-8.2.0-20200404.0-ppc64le-dvd1.iso> -vadapter
vhost2
```

Physical device

A physical USB device can be assigned to a client LPAR to install an operating system by going to the HMC Enhanced GUI and completing the following steps:

1. Select **All Systems**.
2. Click the managed system name.
3. Select **Partitions**.
4. Select **Virtual Storage**.
5. Click the client LPAR name.
6. Select **Physical I/O Adapters**.
7. Click **Add Adapter**.
8. Select the USB adapter.
9. Click **OK**.

Figure 4-32 shows the physical USB adapter assignment to the client LPAR.

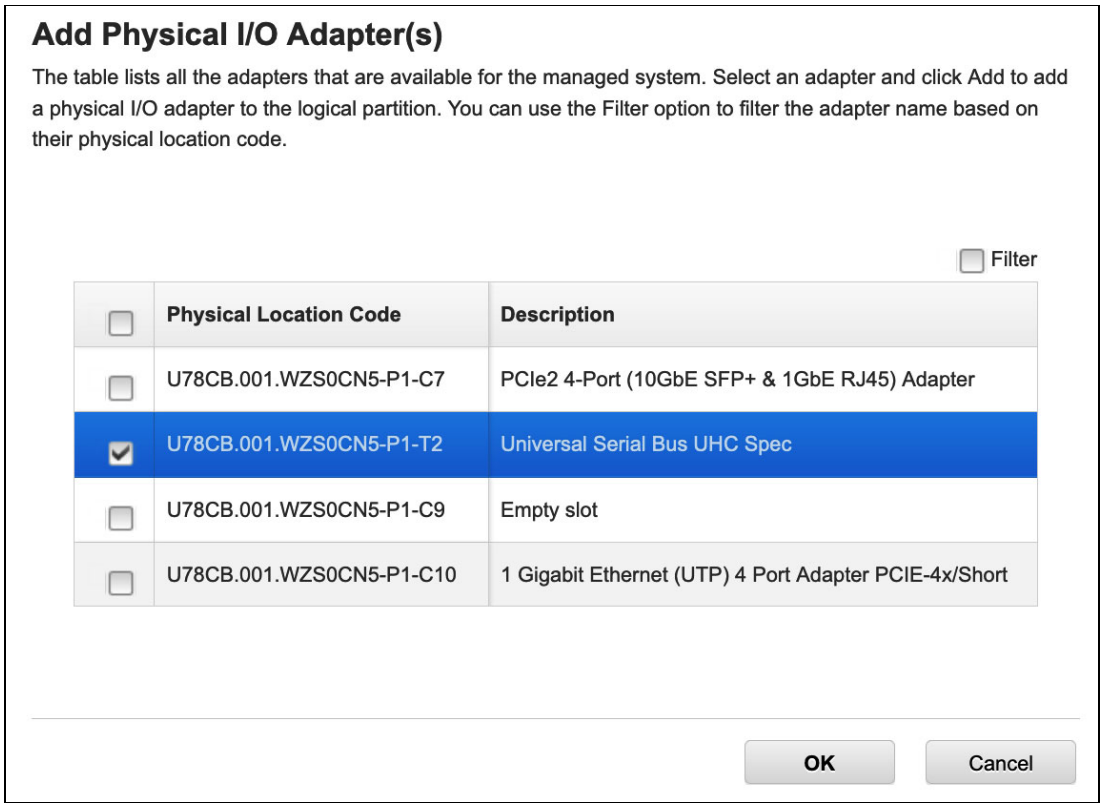


Figure 4-32 Physical USB adapter assignment

Network installation

For network installation on AIX, a NIM can be used, which requires completing the following steps:

1. Download the AIX installation media (ISO files) from the IBM ESS website, found at:

   https://www.ibm.com/servers/eserver/ess/

2. Define `lpp_source` and SPOT. For more information, see How to create a spot and lpp_source from an ISO image, found at:

   https://www.ibm.com/support/pages/node/6485567

3. Allocate resources for network boot (`nim_bosinst`). See Installing a client using NIM, found at:

   https://www.ibm.com/docs/en/aix/7.3?topic=aix-installing-client-using-nim

For IBM i network installation, see IBM i Network Installation Using HMC, found at:

https://www.ibm.com/support/pages/ibm-i-network-installation-using-hmc

For Linux network boot installation on Power, see Installing Red Hat Enterprise Linux on IBM Power System servers using network boot, found at:

https://www.ibm.com/docs/en/linux-on-systems?topic=qsglpss-installing-red-hat-enterprise-linux-power-system-power9-servers-using-network-boot

There are several more methods that you can use to install Linux on your system. For more information, see Additional installation methods, found at:

https://www.ibm.com/docs/en/linux-on-systems?topic=servers-additional-installation-methods

Operating system installation

To install AIX, complete the following steps:

1. Activate the client LPAR in SMS boot mode.
2. Open a console session to the LPAR.
3. Select `5. Select Boot Options` and then press Enter.
4. Select `1. Select Install/Boot Device` and then press Enter.
5. Select `7. List all Devices, look for the virtual CD-ROM/USB` (press `n` to get to the next page if required). Enter the number for the CD-ROM/USB and then press Enter.
6. Select `2. Normal Mode Boot` and then press Enter.
7. Confirm your choice by selecting `1. Yes` and then pressing Enter. ?
8. Next, you are prompted to accept the terminal as `console` and select the installation language.
9. You are presented with the installation menu. It is a best practice to check the settings (option 2) before you proceed with the installation. Check whether the selected installation disk is correct. ?
10. When the installation procedure finishes, use the **`padmin`** username to log in. On initial login, you are prompted to supply the password. There is no default password.
11. After successful login, you are placed under the VIOS CLI.
12. Enter `a` and press Enter to accept the Software Maintenance Agreement terms, then run the following command to accept the license:

```
$ license -accept
```

The AIX installation is complete.

For more information about a Linux installation, see Quick start guides for Linux on IBM Power System servers, found at:

https://www.ibm.com/docs/en/linux-on-systems?topic=systems-quick-start-guides-linux-power-system-servers

For more information about an IBM i installation, see Preparing to install the IBM i release, found at:

https://www.ibm.com/docs/en/i/7.5?topic=partition-preparing-install-i-release

4.5 VIOS security implementation

VIOS offers a set of options to tighten security controls in your VIOS environment.

Through these options, you can select a level of system security hardening and specify the settings that are allowed within that level. With the VIOS security features, you can control network traffic by enabling the VIOS firewall.

The system security hardening feature protects all elements of a system by tightening security or implementing a higher level of security. Although hundreds of security configurations are possible with the VIOS security settings, you can easily implement security controls by specifying a high, medium, or low security level.

You can edit the following security attributes with the system security hardening features that are provided by VIOS:

- Password policy and complexity settings
- System check actions, such as **`usrck`**, **`pwdchk`**, **`grpck`**, and **`sysck`**
- Role-based access control configuration (RBAC)
- Trusted execution and intrusion detection
- Firewall and IP filtering

For more information, see Security on the Virtual I/O Server, found at:

https://www.ibm.com/docs/en/power-sys-solutions/0008-DEA?topic=P8DEA/p8hb1/p8hb1_security.htm

4.5.1 VIOS user types and role-based access control configuration

The VIOS operating system is based on the AIX kernel with a customization to serve the I/O operations to client LPARs.

The first user to log in to VIOS is `padmin` (prime administrator), which is the only active user type when the VIOS is installed.

The prime administrator can create more user IDs with types of system administrator, service representative, development engineer, or other users with different roles. You cannot create the prime administrator (`padmin`) user ID. It is automatically created and enabled, and the role PAdmin is assigned as the default role after the VIOS is installed.

For more information, see Managing users on the Virtual I/O Server, found at:

https://www.ibm.com/docs/en/power10/9105-42A?topic=security-managing-users

A root shell can be access with the **`oem_setup_env`** command for any AIX administration issues on the VIOS operating system. However, the virtualization tasks are done only by the `padmin` user that has access to the virtualization libraries.

You can use RBAC to define roles for users in the VIOS. A role confers a set of permissions or authorizations to the assigned user. Thus, a user can perform only a specific set of system functions that depend on the access rights that are given. For example, if the system administrator creates the role for user management with authorization to access user management commands and assigns this role to a user, that user can manage users on the system but has no further access rights.

For more information, see the following resources:

- Using role-based access control with the Virtual I/O Server, found at:

  https://www.ibm.com/docs/en/power-sys-solutions/0008-DEA?topic=P8DEA/p8hb1/p8hb1_vios_using_rbac.htm

- How to use RBAC on VIOS, found at:

  https://www.ibm.com/support/pages/how-use-rbac-vios

4.5.2 Configuring security hardening (viosecure)

You can configure VIOS security options with the **viosecure** command. To help you set up system security when you initially install the VIOS, VIOS provides the configuration assistance menu. You can access the configuration assistance menu by running the **cfgassist** command.

With the **viosecure** command, you can set, change, and view current security settings. By default, no VIOS security levels are set. You must run the **viosecure** command to change the settings.

For more information, see the viosecure command, found at:

https://www.ibm.com/docs/en/power10/9105-22B?topic=commands-viosecure-command

viosecure also configures, unconfigures, and displays the firewall settings of the network. You can use the **viosecure** command to activate and deactivate specific ports and specify the interface and IP address of the connection. You also can specify to use the IPv6 version of the **viosecure** command to configure, unconfigure, and display the firewall settings of the IPv6 network.

For more information about using **viosecure** for setting up a firewall, see PowerVM: How to use viosecure firewall to deny access to a service for all except for specific IP?, found at:

https://www.ibm.com/support/pages/node/6603069

viosecure can activate, deactivate, and display security hardening rules. By default, none of the security strengthening features are activated after installation. The **viosecure** command guides the user through the proper security settings, which can be high, medium, or low. After this initial selection, a menu is displayed that itemizes the security configuration options that are associated with the selected security level in sets of 10. These options can be accepted in whole, individually toggled, or ignored. After any changes, **viosecure** continues to apply the security settings to the computer system.

For more information, see IBM VIOS: How to create custom viosecure rules, found at:

https://www.ibm.com/support/pages/ibm-vios-how-create-custom-viosecure-rules

VIOS security benchmark

VIOS security benchmark profiles can be applied by using a custom configuration that is part of the **aixpert** tool command, which can be integrated with PowerSC Security and Compliance features.

The Center for Internet Security (CIS) develops benchmarks for the secure configuration of a target system. CIS benchmarks are consensus-based, best-practice, security-configuration guides that are developed and accepted by business and industry.

The CIS specifications for VIOS server provide guidance for establishing a secure configuration by applying the new profiles.

For more information, see CIS specifications for VIOS server, found at:

https://www.ibm.com/docs/en/powersc-standard/2.1?topic=concepts-cis-specifications-vios-server

Several customers are looking for security and compliance automation, which can be achieved with IBM PowerSC.

The PowerSC Security and Compliance Automation feature is an automated method to configure and audit systems in accordance with the US Department of Defense (DoD) Security Technical Implementation Guide (STIG), the Payment Card Industry (PCI) Data Security Standard (DSS), the Sarbanes-Oxley act, COBIT compliance (SOX/COBIT), the Health Insurance Portability and Accountability Act (HIPAA), CIS benchmarks compliance for AIX, and IBM i best practices.

PowerSC helps to automate the configuration and monitoring of systems that must be compliant with the PCI DSS 3.2. Therefore, the PowerSC Security and Compliance Automation feature is an accurate and complete method of security configuration automation that is used to meet the IT compliance requirements of the DoD UNIX STIG, the PCI DSS, t SOX/COBIT, and the HIPAA.

CIS benchmark guidelines are not maintained or supported by IBM; they are directly supported by CIS.

For more information, see Security and Compliance Automation concepts, found at:

https://www.ibm.com/docs/en/powersc-standard/2.1?topic=automation-security-compliance-concepts

4.6 Shared processor pools

Shared processors are described in 2.1.3, “Shared processors” on page 34 and 2.1.5, “Multiple shared processor pools” on page 36. For planning considerations, see 3.2.4, “Shared processor pools capacity planning” on page 82.

This section goes through the shared processor pools (SPPs) feature in PowerVM for allocating and controlling the right capacities.

The default SPP is preconfigured, so you cannot change the properties of the default SPP. The maximum number of processors that are available to the default SPP is the total number of active, licensed processors on the managed system minus the number of processors that are assigned to dedicated processor partitions.

Without using extra SPPs, unused processor capacity is divided among all uncapped LPARs according to their weights within the default SPP. When more SPPs are used, the distribution takes place in two stages. Unused processor shares are first distributed to uncapped LPARs within the same SPP. Only the unused processor shares that are not consumed by other LPARs in the same SPP are redistributed to LPARs in other SPPs.

For more information, see *P*rocessor resource assignment in partition profiles, found at:

https://www.ibm.com/docs/en/power10/9786-42H?topic=profile-processor-resource-assignment

Each SPP has an Entitled Pool Capacity (EPC), which is the sum of the guaranteed entitlements of the assigned LPARs and the Reserved Pool Capacity (RPC). The RPC can be configured by using the **reserved_pool_proc_units** attribute of the SPP and has the default value 0. Just as the entitlement is guaranteed for a shared processor LPAR, the assignment of the EPC is guaranteed for an SPP, regardless of how the shares are distributed to the associated LPARs in the SPP.

For more information, see Changing a shared processor pool, found at:

https://www.ibm.com/docs/en/power10/9786-42H?topic=template-changing-shared-processor-pool-settings

4.7 Active Memory Expansion implementation

Section 2.2.2, “Active Memory Expansion” on page 39 provides an overview of Active Memory Expansion (AME). Section 3.3.2, “Active Memory Expansion planning” on page 90 includes planning considerations.

This section describes how AME can be configured and implemented.

4.7.1 Activating AME

AME can be enabled through the HMC or the PowerVC dashboard. Before you enable AME, verify that your server supports AME (that is, it is AME-capable), as shown in Figure 4-33.

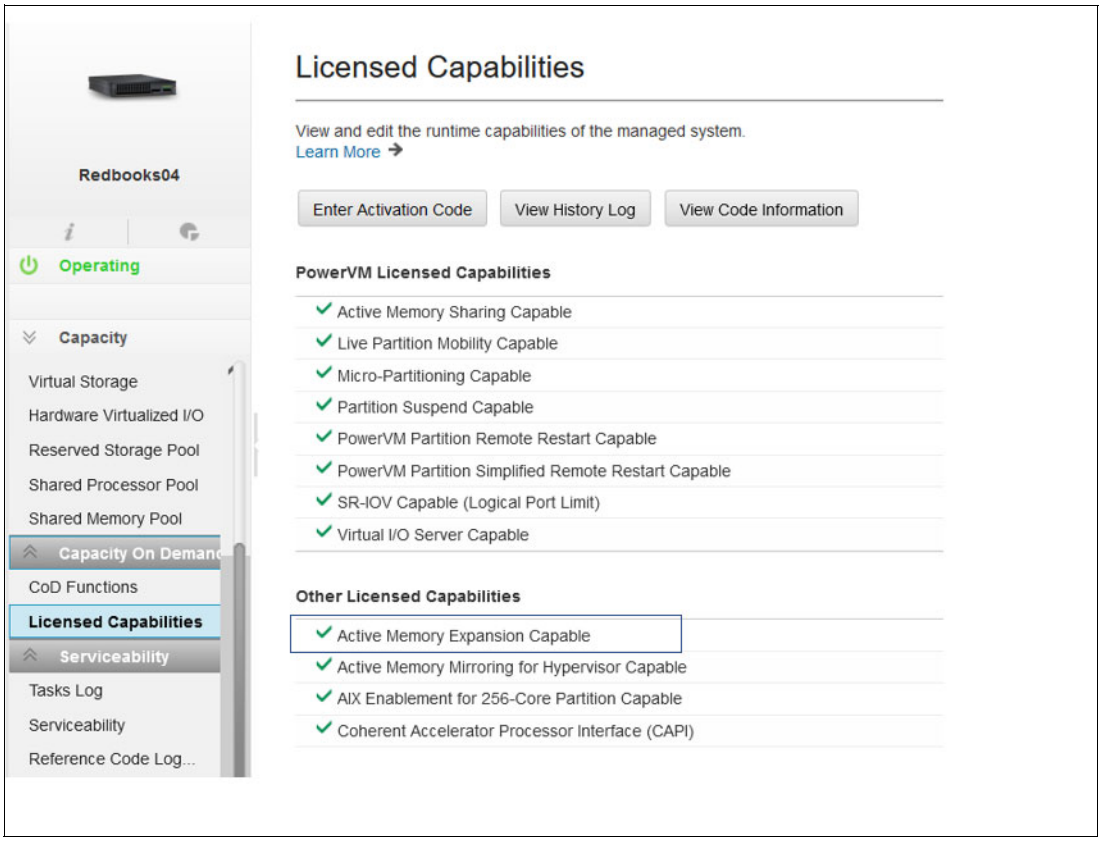


Figure 4-33 AME support on the server

After AME is activated for a server, the feature can be enabled for multiple LPARs that are hosted on that server.

Figure 4-34 summarizes the steps to enable AME for an LPAR.

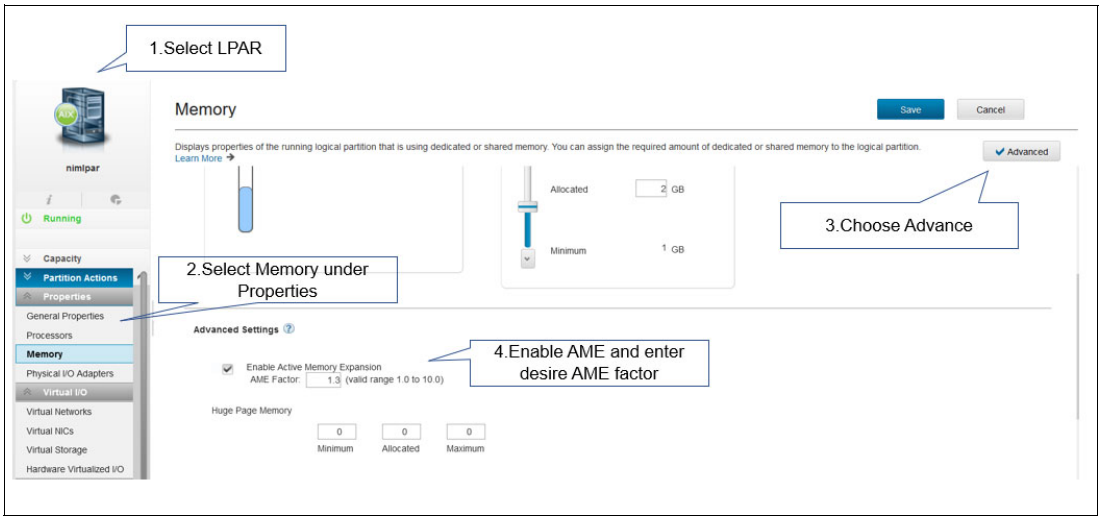


Figure 4-34 Enabling AME for LPAR

Choose the correct Expansion Factor when AME is enabled.

The AME planning tool (**amepat**) that is provided with AIX helps to analyze the workload and provides suggestions for reasonable physical memory size and the respective memory expansion factor. Figure 4-35 shows the **amepat** command output.

```
.
.[Lines omitted for clarity]
.
Active Memory Expansion Modeled Statistics        :
-----------------------------------------
Modeled Expanded Memory Size :  10.00 GB
Achievable Compression ratio   :2.85

Expansion    Modeled True      Modeled             CPU Usage
Factor       Memory Size       Memory Gain         Estimate
---------    -------------     ------------------  -----------
     1.03          9.75 GB       256.00 MB [  3%]   0.00 [  0%]
     1.22          8.25 GB         1.75 GB [ 21%]   0.00 [  0%]
     1.38          7.25 GB         2.75 GB [ 38%]   0.00 [  0%]
     1.54          6.50 GB         3.50 GB [ 54%]   0.00 [  0%]
     1.67          6.00 GB         4.00 GB [ 67%]   0.00 [  0%]
     1.82          5.50 GB         4.50 GB [ 82%]   0.00 [  0%]
     2.00          5.00 GB         5.00 GB [100%]   0.52 [ 26%]
.
.[Lines omitted for clarity]
.
```

Figure 4-35 The amepat command output

In this output, the optimum memory size is 5.5 GB with a memory expansion factor of 1.82. With these settings, the operating system in the partition still sees 10 GB of available memory, but the amount of physical memory can be reduced by almost half. A higher expansion factor means that more CPU resources are needed to perform the compression and decompression. Therefore, choosing the higher value of expansion factor might lead to a higher memory deficit.

For more information, see the **amepat** command, found at:

https://www.ibm.com/docs/en/aix/7.3?topic=amepat-command

For an AME-enabled LPAR, monitoring capabilities are available in standard AIX performance tools, such as **lparstat**, **vmstat**, **topas**, and **svmon**. **amepat** is included with AIX, which enables you to sample workloads and estimate how expandable the partition's memory is, and how many CPU resources are needed.

For more information, see Active Memory Expansion (AME), found at:

https://www.ibm.com/docs/en/aix/7.3?topic=management-active-memory-expansion-ame

Temporary activation of AME

AME is a chargeable feature, which can be purchased separately. You can evaluate the usage of AME at no charge with Trial Capacity on Demand (Trial CoD). With Trial CoD, the AME function can be temporarily activated for up to 60 days at no charge. Trial AME is available once per server and it allows users to validate the benefits that your server can realize.

CoD is described in 2.9, “Capacity on Demand” on page 66.

For more information, see Other Capacity on Demand Advanced Functions, found at:

https://www.ibm.com/docs/en/power10/9105-41B?topic=demand-other-capacity-advanced-functions

4.8 Active Memory Mirroring implementation

Active Memory Mirroring (AMM) is described in 1.3.19, “Active Memory Mirroring” on page 21.

Enabling AMM for the hypervisor doubles the amount of memory that is used by the hypervisor, so available memory for LPARs in the system is affected.

The IBM System Planning Tool (SPT) can provide the estimated amount of memory that is used by the hypervisor. This information is useful when changes are made to an existing configuration or when new servers are deployed.

For more information, see IBM System Planning Tool for Power processor-based systems, found at:

https://www.ibm.com/support/pages/ibm-system-planning-tool-power-processor-based-systems-0

Figure 4-36 highlights the options that must be selected to estimate the amount of memory that is required by the AMM feature.

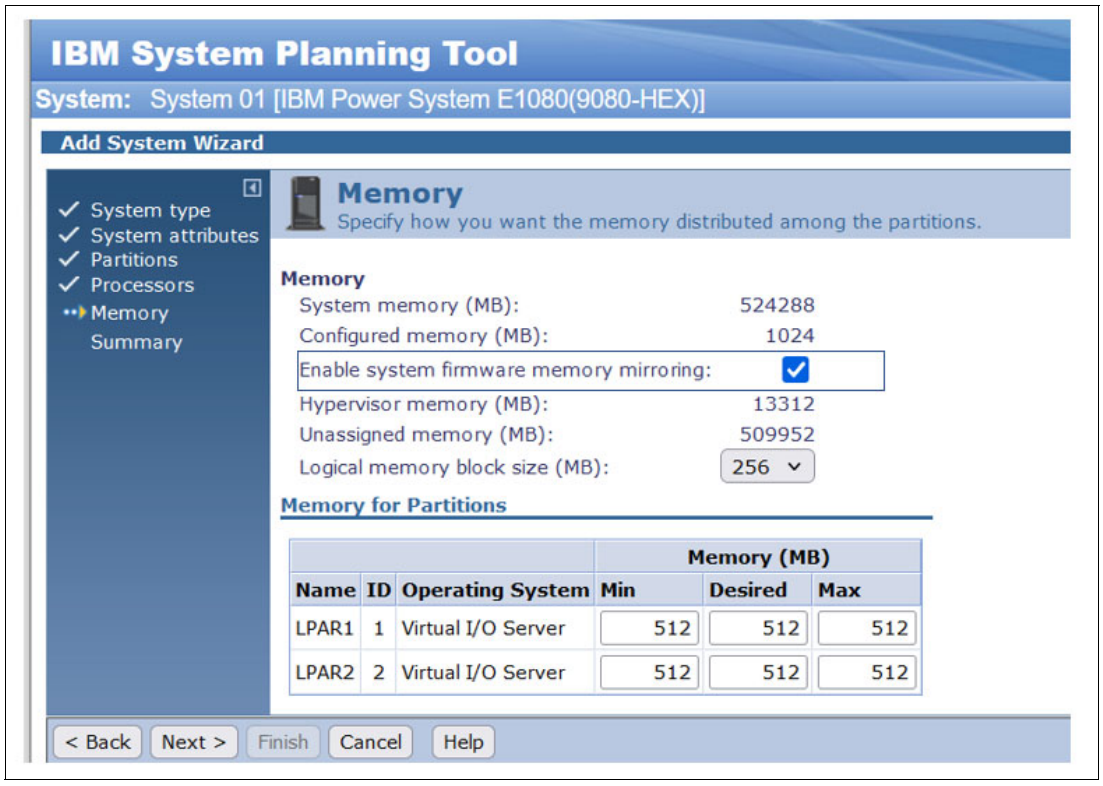


Figure 4-36 Estimating AMM usage by using IBM System Planning Tool

Besides the hypervisor code itself, other components that are vital to the server operation also are mirrored:

- Hardware page tables (HPTs), which are responsible for tracking the state of the memory pages that are assigned to partitions.
- Translation Control Entries (TCEs), which are responsible for providing I/O buffers for the partition's communications.
- Memory that is used by the hypervisor to maintain partition configuration, I/O states, virtual I/O information, and partition state.

It is possible to check whether the AMM option is enabled and change its status by using the HMC. The relevant information and controls are in the Memory Mirroring section of the General Settings window of the selected Power server, as shown in Figure 4-37 on page 177.

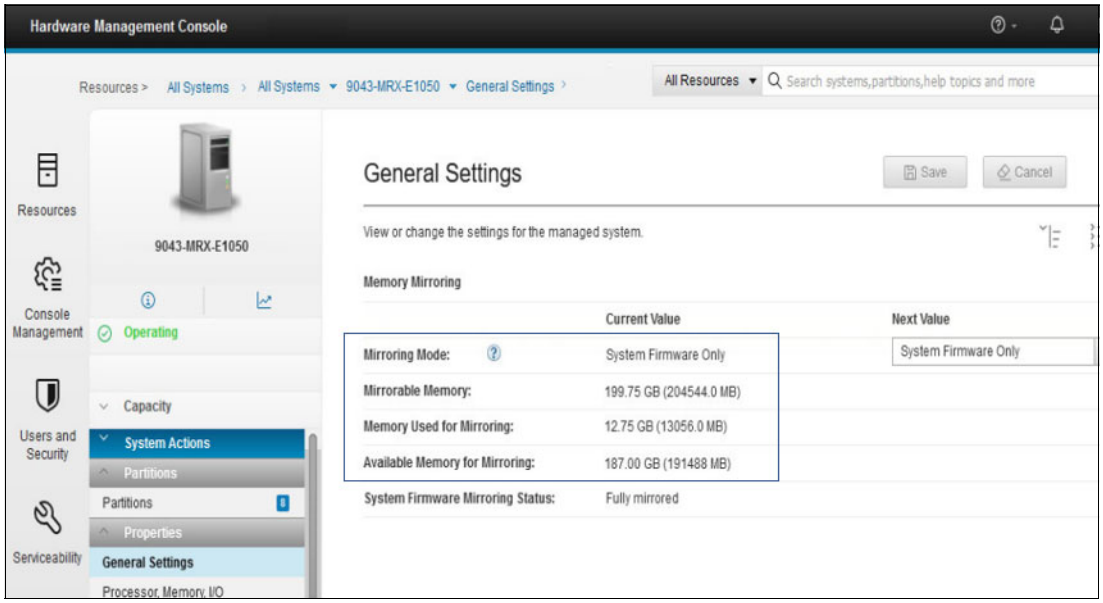


Figure 4-37 Memory Mirroring section in the General Settings window on the HMC GUI

If one of the DDIMMs that contains hypervisor data fails, all the server operations remain active and the eBMC service processor isolates the failing DDIMMs. The system stays in the partially mirrored state until the failing DDIMM is replaced.

Memory that is used to hold the contents of platform dumps is mirrored. AMM does not mirror partition data either. It mirrors only the hypervisor code and its components to protect this data against DDIMMs failures. With AMM, uncorrectable errors in data that is owned by a partition or application are handled by the existing Special Uncorrectable Error (SUE) handling methods in the hardware, firmware, and operating system.

SUE handling prevents an uncorrectable error in memory or cache from immediately causing the system to stop. Rather, the system tags the data and determines whether it is ever used again. If the error is irrelevant, it does not force a checkstop. If the data is used, termination can be limited to the program, kernel, or hypervisor that owns the data, or freeze of the I/O adapters that are controlled by an I/O hub controller if data must be transferred to an I/O device.

All Power10 processor-based enterprise and scale-out servers (except the Power S1014) support the AMM feature.

For more information, see the following IBM Redpapers:

- *IBM Power S1014, S1022s, S1022, and S1024 Technical Overview and Introduction*, REDP-5675
- *IBM Power E1050: Technical Overview and Introduction*, REDP-5684
- *IBM Power E1080 Technical Overview and Introduction*, REDP-5649

4.9 Live Partition Mobility implementation

LPM is described in 2.6.1, “Live Partition Mobility” on page 54.

You must verify that the source and destination systems are configured correctly so that you can successfully migrate the partition. This verification must cover the configuration of the source and destination servers, the HMC, the VIOS LPARs, the partition, the virtual storage configuration, and the virtual network configuration.

For more information about this procedure, see Preparing for partition mobility, found at:

https://www.ibm.com/docs/en/power10/9105-22B?topic=mobility-preparing-partition

The FLRT Live Partition Mobility (LPM) report provides recommendations for LPM operations based on source and target input values. These recommendations might include recommended fixes, including interim fixes, for known LPM issues.

For more information about this tool, see Live Partition Mobility Recommendations, found at:

https://esupport.ibm.com/customercare/flrt/lpm

For a list of LPM best practices, see Best Practices for Live Partition Mobility (LPM) Networking, found at:

https://www.ibm.com/support/pages/best-practices-live-partition-mobility-lpm-networking

For a description of the most common causes that impact LPM performance and considerations to resolve the problem, see Live Partition Mobility Performance, found at:

https://www.ibm.com/support/pages/live-partition-mobility-performance

For more information about how to enable NPIV LUN or disk-level validation on a VIOS for a partition mobility environment, see How to Enable/Disable NPIV LUN or Disk Level Validation on a Virtual I/O Server (VIOS) for Partition Mobility Environment, found at:

https://www.ibm.com/support/pages/how-enabledisable-npiv-lun-or-disk-level-validation-virtual-io-server-vios-partition-mobility-environment

4.10 PowerVC Implementation

Before you can install and use PowerVC, you must ensure that your environment is configured correctly. The environment must be using supported hardware and software with storage, hosts, and network resources configured.

The tasks that are involved in setting up your environment vary depending on whether you are installing PowerVC in an existing environment or in a new environment. For new environments, you might be using new hardware or repurposing existing hardware to create your environment.

For more information about planning and configuration, see Setting up the PowerVC environment, found at:

https://www.ibm.com/docs/en/powervc/2.1.0?topic=setting-up-powervc-environment

PowerVC can be deployed on a virtual machine (VM), and as a best practice, PowerVC should be the only application on that VM. However, PowerVC can generally coexist with other software on the same instance, assuming that there is no resource or dependency conflict between PowerVC and the other software. Potential conflicts include, for example, port contention, user namespace, file system capacity, and firewall settings.

You must consider performance implications to PowerVC and the other software when you install other software on the same instance. For example, PowerVC memory usage might grow and cause problems with applications that coexist with PowerVC. PowerVC resource requirements are sized by assuming that PowerVC is the only workload that is running on the management instance. If other applications are using resources, adjust the sizing as required.

Consider the following points before you start the PowerVC installation procedure:

- Make sure that the managed hosts are on an IBM Power8 processor-based server or later.
- If you have any previous version of PowerVC that is installed, take a backup, copy the backup file to a custom location, and then uninstall the existing version. Restart the system and install the new version of PowerVC.
- Review the hardware and software requirements.
- For RHEL and SUSE Linux Enterprise Server, a PowerVC 2.1.0 installation is supported on both single-node and multinode environments.
- Make sure that you disable IPv6 before you proceed with the installation procedures.

To install PowerVC, complete the following steps:

1. Configure these repositories for RHEL and SUSE Linux Enterprise Server as based on the environment:
   - A YUM repository for PowerVC that is installed through RHN. Make sure that the following repositories are enabled.
     - AppStream
     - BaseOS
     - Supplementary
     - HA
     - Ansible
   - A Zypper repository for PowerVC that is installed through SUSE Linux Enterprise Server 15 SP2 and SUSE Linux Enterprise Server 15 SP3. Make sure that the following repositories are enabled.
     - `SLE-Module-Basesystem`
     - `SLE-Module-Desktop-Applications`
     - `SLE-Module-Development-Tools`
     - `SLE-Module-Legacy`
     - `SLE-Module-Public-Cloud`
     - `SLE-Module-Server-Applications`
     - `SLE-Module-Web-Scripting`
     - `SLE-Product-HA`
     - `SLE-Product-SLES`

2. Extract the compressed file that matches your environment to the location from which you want to run the installation script:
   - For ppc64le, extract `download_location/powervc-opsmgr-<rhel or sles>-ppcle-<powervc_version>.tgz`, where `download_location` is the directory to which the file was downloaded.
   - For x86_64, extract `dvd_mount_point/powervc-opsmgr-rhel-x86-<powervc_version>.tgz`, where `dvd_mount_point` is the directory where the ISO image was mounted.
3. Change the directory to `extract location/powervc-opsmgr-<version>`, where the extraction location is the directory that you extracted.
4. After the installation is complete, access IBM Fix Central to download and install any fix packs that are available. For more information, see IBM Fix Central, found at:

   https://www.ibm.com/support/fixcentral/

After you install PowerVC, you can add, configure, and manage the resources. For more information, see Adding resources to PowerVC, found at:

https://www.ibm.com/docs/en/powervc/2.1.0?topic=configuring-adding-resources-powervc

After you install PowerVC, you can access it by opening your browser and entering the URL `https://`*`powervc_hostname`* or `https://`*`powervc_IP_address`*. Then, you can log in to PowerVC for the first time with the root credentials of the management host where PowerVC is installed.

Note: To register resources, you must sign in with the credentials of a user with the admin role.

For more information, see Getting started with PowerVC, found at:

https://www.ibm.com/docs/en/powervc/2.1.0?topic=getting-started-powervc

5

Managing the IBM PowerVM environment

This chapter provides guidelines to manage your PowerVM environment.

The IBM PowerVM environment is composed of multiple components like an IBM Power server, PowerVM hypervisor (PHYP), Virtual I/O Server (VIOS), client logical partitions (LPARs), and many other capabilities that PowerVM offers. Therefore, management of a PowerVM environment consists of managing all these components. This chapter describes management, best practices, and guidelines to manage your PowerVM environment.

The chapter also covers important aspects of management like monitoring LPARs. It also provides an overview of solutions that are available on PowerVM.

The tools that are available to manage Power servers are also described.

This chapter covers the following topics:

- Hardware Management Console management best practices
- Firmware management best practices
- VIOS management best practices
- LPAR management best practices
- Management solutions on PowerVM
- Management tools on Power servers
- PowerVC

5.1 Hardware Management Console management best practices

The Hardware Management Console (HMC) provides a simplified interface to manage Power servers.

For a description of the menu options and tasks that are available in the HMC, see Overview of menu options, found at:

https://www.ibm.com/docs/en/power10/7063-CR2?topic=hmc-overview-menu-options

Figure 5-1 highlights the HMC Management options.

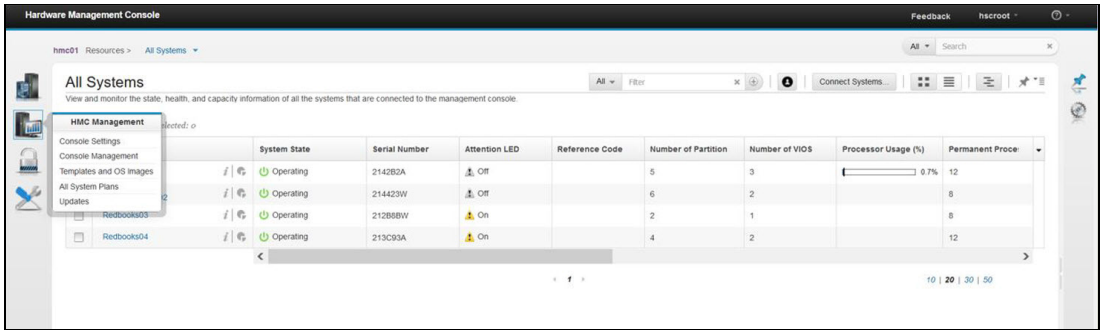


Figure 5-1 HMC Management

As a best practice, configure dual HMCs or redundant HMCs for managing your Power servers. When two HMCs manage one system, they are peers, and each HMC can be used to control the managed system.

For more information about configuration guidelines, see Hardware Management Console virtual appliance, found at:

https://www.ibm.com/docs/en/power10/000V-HMC

5.1.1 HMC upgrades

Updates and upgrades are released periodically for the HMC. As part these updates, new functions and improvements are added to the HMC.

For more information about maintaining your HMC, see Updating, upgrading, and migrating your HMC machine code, found at:

https://www.ibm.com/docs/en/power10/7063-CR1?topic=hmc-updating-upgrading-migrating-your-machine-code

5.1.2 HMC backup and restore

As a best practice, back up the HMC data after any changes are made to the HMC or to the information that is associated with LPARs.

The HMC data that is stored on the HMC hard disk drive can be saved to a DVD on a local file system, a remote system that is mounted to the HMC file system (such as NFS), or sent to a remote site by using File Transfer Protocol (FTP).

Figure 5-2 highlights the options that are available on the HMC GUI to back up and restore the HMC.

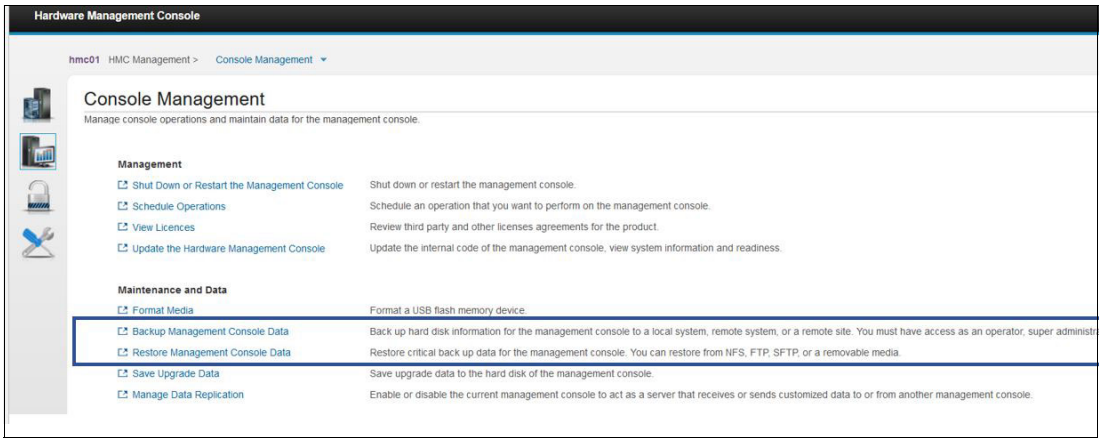


Figure 5-2 HMC backup and restore

To learn about the tasks that are available on the HMC under HMC Management, see Console Management tasks, found at:

https://www.ibm.com/docs/en/power10/7063-CR2?topic=earlier-console-management-tasks

5.1.3 HMC monitoring capabilities

The HMC plays a crucial role in the management of servers. Several tasks are run from the HMC to manage VIOS partitions and client LPARs, servers, and resources that are available on the servers.

HMC captures logs and service events, which enable the system administrator to view the tasks that are performed and analyze the events that are generated. The HMC also plays an important role in the Call Home function. If Call Home is enabled for a server that is managed by HMC, the console automatically contacts a service center when a serviceable event occurs. When the Call Home function on a managed system is disabled, your service representative is not informed of serviceable events.

HMC also provides connection monitoring capabilities. Connection monitoring generates serviceable events when communication problems are detected between the HMC and the managed systems. If you disable connection monitoring, no serviceable events are generated for networking problems between the selected machine and this HMC.

For a description of the tasks that are available on the HMC for the Serviceability tasks, see Serviceability tasks, found at:

https://www.ibm.com/docs/en/power10/7063-CR1?topic=earlier-serviceability-tasks

5.2 Firmware management best practices

Server firmware is the code that is in system flash memory. It includes several subcomponents, such as PHYP, power control, the service processor, and LPAR firmware that is loaded into either AIX, IBM i, or Linux LPARs.

Firmware management plays a crucial role in the overall Power server management strategy.

Firmware updates are used for:

- Adding functions to an existing system or supporting a new system. These updates are added as part of a major release.
- Providing fixes or group of fixes to the existing release.

5.2.1 Firmware terminology

This section introduces terms that are used in the context of firmware management.

- Release level

  A major new function, such as the introduction of new hardware models and functions and features that are enabled through firmware. This firmware upgrade is disruptive.

- Service Pack (SP)

  Primarily firmware fixes and minor function changes that are applicable to a specific release level. These firmware updates are usually concurrent.

- Types of firmware updates:
  - Concurrent

    A code update that allows the operating systems that run on the Power server to continue running while the update is installed and activated.

  - Deferred

    A code fix that is concurrent but not activated on the system until the Power server is restarted.

  - Partition deferred

    A code fix that is concurrent but not activated until the partition is reactivated.

  - Disruptive

    A code fix, which requires a Power server restart during the code update process.

Note: Deferred, partition-deferred, and disruptive content is identified in the firmware readme file.

5.2.2 Determining the type of firmware installation

The information in this section helps you to determine whether your installation will be concurrent or disruptive.

For systems that are not managed by an HMC, the installation of system firmware is always disruptive.

Note: The file names and SP levels that are used in the following examples are for clarification only. They are not necessarily levels that were released in the past or will be released in the future.

The naming convention for system firmware files is as follows:

- Example: 01VH*xxx_yyy_zzz*
  - xxx is the release level.
  - yyy is the SP level for xxx release level.
  - zzz is the last disruptive SP level for xxx release level.

Note: Values of SP level and last disruptive SP level (yyy and zzz) are only unique within a release level (xxx). For example, 01VH900_040_040 and 01VH910_040_045 are different SPs.

An installation is disruptive in the following situations:

- The release levels (xxx) are different. For example, the currently installed release is 01VH900_040_040. The new release is 01VH910_040_040.
- The SP level (yyy) and the last disruptive SP level (zzz) are the same. For example, VH910_040_040 is disruptive, no matter what level of VH910 is installed on the system.
- The SP level (yyy) that is installed on the system is earlier than the last disruptive SP level (zzz) of the SP to be installed. For example, the currently installed SP is VH910_040_040 and the new SP is VH910_050_045.

An installation is concurrent if the release level (xxx) is the same and the SP level (yyy) that is installed on the system is the same or later than the last disruptive SP level (zzz) of the SP to be installed. For example, the currently installed SP is VH910_040_040, and the new SP is VH910_041_040.

The definitions of impact area and SP severity that are provided with the firmware details are listed in Preventive Service Planning, found at:

https://www.ibm.com/support/pages/glossary

For more information about how to view the current firmware level that is installed on an AIX, Linux, and IBM i partition by using Advanced System Management Interface (ASMI) or HMC, see Viewing existing firmware levels, found at:

https://www.ibm.com/docs/en/power10/9786-22H?topic=fixes-viewing-existing-firmware-levels

5.2.3 Planning firmware updates and upgrades

As a best practice, plan for firmware maintenance twice a year, although the frequency can be tailored to suit the customer's environment. If firmware maintenance is planned twice a year, one of the maintenance windows can support an upgrade to move off a release that is no longer supported. The other maintenance window can be for an update, that is, move to a newer SP at the current release, which can be done concurrently.

5.2.4 System firmware maintenance best practices

Here are best practices to follow for system firmware updates and upgrades:

- When a new SP is released, review the readme file for the SP. Check the SP Fix List to see whether any critical fixes are applicable to your environment and configuration.
- If an SP includes a High Impact PERvasive (HIPER) fix that is applicable to your environment, install the SP as soon as a maintenance window can be scheduled.
- A “deferred fix” may remain pending until the next scheduled restart.
- Use a Release Level that is supported by the SPs.
- If you do not require the features and functions that are introduced by a new Release Level, you might stay on the older Release Level if it is still supported. SPs continue to be delivered for supported Release Levels.

Server Firmware Update and Upgrade Instructions, found at https://www.ibm.com/support/pages/server-firmware-update-and-upgrade-instructions, includes information about the following firmware maintenance tasks:

- For HMC-managed systems, how to concurrently update server firmware.
- For HMC-managed systems, how to disruptively update server firmware.
- For IBM i stand-alone systems (non-HMC-managed systems).

For more information about updating the firmware on a system that is managed by only PowerVM NovaLink, see Updating the firmware on a system that is managed by PowerVM NovaLink, found at:

https://www.ibm.com/docs/en/power10/9105-22A?topic=upn-updating-firmware-system-that-is-managed-by-powervm-novalink

As described in Chapter 1, “IBM PowerVM overview” on page 1, Power10 processor-based scale-out and midrange servers use the enterprise Baseboard Management Controller (eBMC) service processor instead of the Flexible Service Processor (FSP).

For eBMC-based servers, you might perform a firmware update directly from the eBMC ASMI menu. For more information, see Firmware update via eBMC ASMI menu, found at:

https://mediacenter.ibm.com/media/Firmware+update+via+eBMC+ASMI+menu/1_ca8z6yff

5.2.5 I/O adapter firmware management

For more information about I/O firmware and I/O firmware updates, see I/O Firmware, found at:

https://www.ibm.com/docs/en/power10/7063-CR1?topic=firmware-io

For more information about single-root I/O virtualization (SR-IOV) firmware or to update the driver firmware for shared SR-IOV adapters, see SR-IOV Firmware, found at:

https://www.ibm.com/docs/en/power10/7063-CR1?topic=firmware-sr-iov

5.3 VIOS management best practices

This section describes VIOS management aspects and best practices in the implementation design.

VIOS facilitates the sharing of physical I/O resources between client LPARs within the server. Because the VIOS serves the client LPARs, it must be maintained and monitored continuously. VIOS has several areas that must be planned and maintained in a production environment, such as the following items:

- VIOS version updates and upgrades
- Design of single, dual, or more VIOS per server
- Live Partition Mobility (LPM) prerequisites for VIOS
- Shared Ethernet Adapters (SEA) high availability (HA) modes
- Virtual adapter tunables
- I/O reporting

The implementation of VIOS in a Power server can take the form of single VIOS managing the virtual I/O or more than a single VIOS for redundancy.

The key here is to understand your availability requirements and decide what you need. If the redundancy is achieved at the hardware level, then redundancy at the application level across multiple LPARs or physical frames might not be needed or vice-versa based on your planning and environment readiness.

5.3.1 Single VIOS

Power servers are resilient by design, so depending on the Power server model and the availability requirements, a single VIOS can be used to support some small and non-production workloads. However, eliminating single points of failure (SPOFs) by implementing dual VIOS is a best practice.

Also, some environments have heavy network and storage I/O traffic. In such cases, consider using multiple VIOSs to isolate network traffic from the I/O from the storage area network (SAN) by using different set of VIOS pairs within the same server.

Overall, using dual or multiple VIOSs per machine are preferred for most situations.

5.3.2 Dual or multiple VIOSs

The benefits of using multiple VIOSs are improved availability and performance.

Mission-critical workloads have different requirements than development and test partitions. Therefore, it is a best practice to plan for the right amount of resources that are needed by one or multiple VIOSs, depending on the workloads that you plan to run.

Installing dual VIOSs in the same Power server is a best practice, and it is the mandatory setup for redundancy. With dual VIOS, you can restart one VIOS for a planned maintenance operation and keep the client LPARs running.

A dual-VIOS setup provides redundancy, accessibility, and serviceability. It offers load-balancing capabilities for multipath input/output (MPIO), multiple SEA, and virtual Network Interface Controllers (vNICs) failover configurations.

Compared to a single VIOS setup, a dual-VIOS setup has the following extra components:

- VIOS pairs communicate with each other by using a control channel over virtual local area network (VLAN) ID 4095 on virtual switch when a simplified SEA is used.
- Setting the trunk priority on the virtual trunk Ethernet adapters that are used in a SEA configuration. The trunk priority determines which VIOS is the primary in a SEA failover setup.

A dual VIOS configuration allows the client LPARs to have multiple paths (two or more) to their resources. In this configuration, if one of the paths is not available, the client LPAR can still access its resources through another path.

These multiple paths can be used to set up HA I/O virtualization configurations, and it can provide multiple ways for building high-performance configurations. These goals are achieved with the help of advanced capabilities that are provided by PowerVM (VIOS and PHYP) and the operating systems on the client LPARs.

Both HMC and NovaLink allow configuration of dual VIOSs on managed systems.

For the storage, PowerVM offers three types of virtualization for client LPARs for enhanced storage availability to client partitions:

- Virtual SCSI (vSCSI)
- N_Port ID Virtualization (NPIV)
- Shared storage pool (SSP)

For more information about multi-pathing configurations, see the following resources:

- Multipathing and disk resiliency with vSCSI in a dual VIOS configuration, found at:

  https://www.ibm.com/support/pages/multipathing-and-disk-resiliency-vscsi-dual-vios-configuration

- Path control module attributes, found at:

  https://www.ibm.com/docs/en/aix/7.3?topic=io-path-control-module-attributes

For more information about dual-VIOS configurations, see the following resources:

- Creating partition profiles for dual VIOS, found at:

  https://www.ibm.com/docs/en/power-sys-solutions/0008-DEA?topic=P8DEA/p8eew/p8eew_create_dual_viosprofile.htm

- Configuring VIOS partitions for a dual setup, found at:

  https://www.ibm.com/docs/en/power-sys-solutions/0008-DEA?topic=P8DEA/p8eew/p8eew_configure_dual_vios.htm

- Virtual Storage Redundancy with dual VIOS Configuration, found at:

  https://community.ibm.com/community/user/power/blogs/robert-kovacs1/2020/07/22/virtual-storage-redundancy-with-dual-vios-configur

5.3.3 VIOS backup and restore

This section describes the PowerVM backup and restore methodology and its importance.

It is important to keep VIOS up to date and backed up because it is a critical part of your infrastructure. When you plan for VIOSs, it is important to plan your VIOS backups. You can back up to a Network Installation Manager (NIM) server, tape, DVD, NFS server, or to IBM Spectrum Storage solutions. Using HMC V9R2M950 or later, you can back up your VIOS I/O configuration and your VIOS image to the HMC and restore them later from the HMC.

Before any updates, you can use the **alt_root_vg** command to clone **rootvg** so that you have a fast failback if VIOS runs into issues. For more information, see How to clone a PowerVM VIOS rootvg?, found at:

https://www.ibm.com/support/pages/how-clone-powervm-vios-rootvg

VIOS backup by using VIOS tools

You can choose either to back up the VIOS operating system as a bootable backup, or back up only the virtual mappings and configuration backup.

VIOS system backup

You can create an installable image of the root volume group by using the **backupios** command. The **backupios** command creates a backup of the VIOS and places it in a file system, bootable tape, or DVD. You can use this backup to reinstall a system to its original state after it was corrupted. If you create the backup on tape, the tape is bootable and includes the installation programs that are needed to install from the backup.

The **backupios** command can use **-cd** flag to create a system backup image to DVD-RAM media. If you must create multi-volume discs because the image does not fit on one disc, the **backupios** command gives instructions for disk replacement and removal until all the volumes are created.

The **backupios** command can use the **-file** flag to create a system backup image to the path that is specified. The file system must be mounted and writable by the VIOS root user before the **backupios** command is run.

For more information, see the **backupios** command, found at:

https://www.ibm.com/docs/en/power10?topic=commands-backupios-command

VIOS configuration backup

Back up the virtual definition configurations regularly. Use the **viosbr** command to back up the virtual and logical configuration, list the configuration, and restore the configuration of the VIOS.

The **viosbr** command backs up all the relevant data to recover VIOS after a new installation. The **viosbr** command has the **-backup** parameter that backs up all the device properties and the virtual devices configuration on the VIOS. This backup includes information about logical devices, such as storage pools, file-backed storage pools, and the virtual media repository.

The backup also includes the virtual devices, such as Etherchannel, SEAs, virtual server adapters, the virtual log repository, and server virtual Fibre Channel (SVFC) adapters. Additionally, it includes the device attributes, such as the attributes for disks, optical devices, tape devices, Fibre Channel (FC) SCSI controllers, Ethernet adapters, Ethernet interfaces, and logical Host Ethernet adapters (HEAs).

The **viosbr** command can run once, or run in a specified period by using the **-frequency** parameter with the daily, weekly, or monthly option. Daily backups occur at 00:00, weekly backups on Sundays at 00:00, and monthly backups on the first day of the month at 00:01. The **-numfile** parameter specifies the number of successive backup files that are saved, with a maximum value of 10. After the specific number of files is reached, the oldest backup file is deleted during the next backup cycle.

Example 5-1 shows a backup of all the device attributes and virtual device mappings daily on the VIOS, keeping the last seven backup files.

Example 5-1 A viosbr backup with frequency options

```
$viosbr –backup -file vio1_backup –frequency daily numfiles 7
```

The backup files that result from running this command are under `/home/padmin/cfgbackups` with the following names for the seven most recent files:

- `vios1_backup.01.tar.gz`
- `vios1_backup.02.tar.gz`
- `vios1_backup.03.tar.gz`
- `vios1_backup.04.tar.gz`
- `vios1_backup.05.tar.gz`
- `vios1_backup.06.tar.gz`
- `vios1_backup.07.tar.gz`

All the configuration information is saved in a compressed XML file. If a location is not specified with the **-file** option, the file is placed in the default location `/home/padmin/cfgbackups`.

For more information, see the **viosbr** command, found at:

https://www.ibm.com/docs/en/power10?topic=commands-viosbr-command

VIOS backup by using HMC

Starting with HMC V9R2M950, a new GUI feature was added to back up a full VIOS (**backupios**) and the VIOS configuration (**viosbr**).

Starting with V10R1M1010, the following commands were added to support VIOS backup and restore:

- **mkviosbk**
- **lsviosbk**
- **rstviosbk**
- **rmviosbk**
- **cpviosbk**
- **chviosbk**

These commands provide the same function through the HMC command-line interface (CLI).

When the backup operation is performed from the HMC, it calls the VIOS to initiate the backup. Then, the backup operation attempts to do a secure copy (**scp**) to transfer the backup from the VIOS to the HMC.

The following prerequisites must be met for this backup operation to complete successfully:

- Resource Monitoring and Control (RMC) must be active for the VIOS that is going to be backed up.
- Enough space in VIOS `/home/padmin` and HMC `/data/`.
- The Secure Shell (SSH) must be working from HMC to VIOS and from VIOS to HMC (port number 22 must be allowed in any firewall between HMC and VIOS).

For more information about managing VIOS backups from HMC, see Manage Virtual I/O Server Backups, found at:

https://www.ibm.com/docs/en/power10?topic=images-manage-virtual-io-server-backups

If you decide to use the HMC CLI for VIOS backups, you must be at HMC V10R1M1010 or later.

To perform VIOS backups from the HMC CLI, you can use one of the following three backup types:

- The **vios** type is for a full VIOS backup. Example 5-2 shows a VIOS full backup from the HMC.

Example 5-2 VIOS full backup from the HMC

```
$ mkviosbk -t vios -m sys1 -p VIOS_NAME -f vios1_full_backup -a
"nimol_resource=1,media_repository=1"
```

- The **viosioconfig** type is for a VIOS I/O configuration backup. Example 5-3 shows a VIOS I/O configuration backup from the HMC.

Example 5-3 VIOS I/O configuration backup

```
$ mkviosbk -t viosioconfig -m sys1 -p VIOS_NAME -f vios1_io_backup
```

- A **ssp** configuration backup. Example 5-4 shows VIOS SSP configuration backup from the HMC.

Example 5-4 VIOS SSP configuration backup

```
$ mkviosbk -t ssp -m sys1 -p vios1 -f vios1_ssp_backup
```

For the syntax of the **mkviosbk** command, see HMC Manual Reference Pages - MKVIOSBK, found at:

https://www.ibm.com/docs/en/power10?topic=commands-mkviosbk

If you decided to restore those VIOS configurations, use the **rstviosbk** command, as shown in Example 5-5.

Example 5-5 Restoring the VIOS configuration by using the rstviosbk command

```
$ rstviosbk -t viosioconfig -m sys1 -p VIOS_NAME -f vios1_io_backup
$ rstviosbk -t ssp -m sys1 -p VIOS_NAME -f vios1_ssp1_backup
```

For more information, see HMC Manual Reference Pages - RSTVIOSBK, found at:

https://www.ibm.com/docs/en/power10?topic=commands-rstviosbk

5.3.4 VIOS upgrade

This section covers the VIOS upgrade and releases notes.

To ensure the reliability, availability, and serviceability (RAS) of a computing environment that uses the VIOS, update the VIOS software to the most recent fix level for that release. The most recent level contains the latest fixes for the specified VIOS release. You can download the most recent updates for VIOS from the IBM Fix Central website, found at:

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

For more information about VIOS releases, see Virtual I/O Server release notes, found at:

https://www.ibm.com/docs/en/power10/9080-HEX?topic=environment-virtual-io-server-release-notes

It is important to keep your VIOS up to date and backed up because it is a critical part of your infrastructure. For more information about current support, see PowerVM Virtual I/O Server on FLRT Lite, found at:

https://esupport.ibm.com/customercare/flrt/liteTable?prodKey=vios

The upgrade steps depend on the level that is installed. For example, to upgrade to VIOS 3, Version 2.2.6.32 or later must be installed.

The base code is downloaded from the IBM Entitled Systems Support (IBM ESS) website, found at:

https://www.ibm.com/servers/eserver/ess/OpenServlet.wss

When you download the code for your entitled software, you see PowerVM Enterprise ED V3. Technology levels and SPs are downloaded from Fix Central. You can download the flash image because it is a fully updated PowerVM 3.1 image.

Upgrade methods

Various ways to upgrade the VIOS to Version 3 are available:

- Use the **viosupgrade** command on the VIOS itself.

  For more information, see the **viosupgrade** command, found at:

  https://www.ibm.com/docs/en/power10/9080-HEX?topic=commands-viosupgrade-command
- Use the **viosupgrade** command with AIX NIM.
- Use the manual upgrade. The manual upgrade includes three steps:
  - Manually back up the VIOS metadata by using the **viosbr -backup** command.
  - Install VIOS 3 through the AIX NIM Server, Flash Storage, or HMC.
  - Restore the VIOS metadata by using the **viosbr -restore** command.

For more information, see the following resources:

- Migrating the Virtual I/O Server by using the **viosupgrade** command or by using the manual method, found at:

  https://www.ibm.com/docs/en/power10/9105-22B?topic=migrating-virtual-io-server-by-using-viosupgrade-command
- Supported Virtual I/O Server upgrade levels, found at:

  https://www.ibm.com/docs/en/power10/9105-22B?topic=command-supported-virtual-io-server-upgrade-levels

Viosupgrade tool considerations

To use the **viosupgrade** command upgrade approach, complete the following steps:

1. Download the latest VIOS flash media.
2. The target VIOS must be at SP 2.
3. Add empty disk to the VIOS for the **alt_clone** command.
4. Mount the `mksysb.image` from the ISO image that you downloaded and uploaded to VIOS, as shown in Example 5-6.

Example 5-6 Mounting mksysb.image from the VIOS ISO image

```
# loopmount -i /tmp/VIOS.iso -o "-V udfs -o ro" -m /mnt
```

5. Copy the `mksysb` file from the mounted image to a different location as shown in Example 5-7.

Example 5-7 Copying the mksysb image to another location

```
# cp /mnt/usr/sys/inst.images/mksysb_image /home/VIO31_mksysb_image
```

6. Start the upgrade process from the copied **mksysb** file as shown in Example 5-8.

Example 5-8 Using the viosupgrade command to start the process

```
# viosupgrade -l -i /home/VIO31_mksysb_image -a hdisk3
```

7. Check the upgrade status by using the **viosupgrade -l -q** command.

For more information, see Upgrading to VIOS 3.1, found at:

https://www.ibm.com/support/pages/upgrading-vios-31

If you are confused about the upgrade path to follow, see Table 5-1 and Table 5-2 on page 194.

Table 5-1 Upgrade path for VIOSs that are not configuring SSP clusters

VIOS level	Procedure to upgrade	Alternative procedure
1.5.x	1. Migrate to 2.1. 2. Update to SP 2.2.6.10. 3. Update to SP 2.2.6.61. 4. Update to SP 2.2.6.65. 5. Upgrade to 3.1.	1. Reinstall at 3.1. 2. Restore a configuration from backup.
2.1 - 2.2.6.0	1. Update to SP 2.2.6.10. 2. Update to SP 2.2.6.61. 3. Update to SP 2.2.6.65. 4. Upgrade to 3.1.	1. Reinstall at 3.1. 2. Restore a configuration from backup.

Table 5-2 Upgrade path for VIOSs that are configuring SSP clusters

VIOS level	Procedure to upgrade	Alternative procedure
2.2.0.11 - 2.2.1.3	1. Update to 2.2.1.4 or 2.2.1.5 2. Update to SP 2.2.6.10 3. Update to SP 2.2.6.61 4. Update to SP 2.2.6.65 5. Upgrade to 3.1	1. Reinstall at 3.1 2. Restore configuration from the backup
2.2.1.4 - 2.2.6.0	1. Update to SP 2.2.6.10 2. Update to SP 2.2.6.61 3. Update to SP 2.2.6.65 4. Upgrade to 3.1	1. Reinstall at 3.1 2. Restore configuration from the backup

5.3.5 VIOS monitoring

As a best practice, monitor the VIOSs in your environment. A basic level of monitoring is to use the error log to monitor different types of errors.

For more information, see the **errlog** command, found at:

https://www.ibm.com/docs/en/power10?topic=commands-errlog-command

While administrators are reviewing the report that is generated by **errlog**, they can use a tool called **summ**, which decodes FC and SCSI disk AIX error report entries. It is an invaluable tool that can aid in diagnosing storage array or SAN fabric-related problems, and it provides the source of the error.

For more information, see the **summ** tool, found at:

https://www.ibm.com/support/pages/node/1072626

You can configure AIX syslog, which provides an extra option for defining different event types and the level of criticality.

For more information, see VIOS syslog, found at:

https://www.ibm.com/support/pages/vios-syslog

An excellent option for VIOS performance monitoring is the **nmon** command. **nmon** can run either in interactive or recording mode. It can display system statistics in interactive mode or to record them in the file system for later analysis.

If you specify any of the **-F**, **-f**, **-X**, **-x**, and **-Z** flags, the **nmon** command is in recording mode. Otherwise, the **nmon** command is in interactive mode.

The **nmon** command provides the following views in interactive mode:

- Adapter I/O statistics (pressing the a key)
- I/O processes view (pressing the A key)
- Detailed PAdapter I/O statistics (pressing the a key)
- I/O processes view (pressing the A key)
- Detailed Page Statistics (pressing the M key)
- Disk busy map (pressing the o key)
- Disk groups (pressing the g key)
- Disk statistics (pressing the D key)

- Disk statistics with graph (pressing the d key)
- IBM ESS vpath statistics view (pressing the e key)
- FC adapter statistics (pressing the ^ key)
- JFS view (pressing the j key)
- Kernel statistics (pressing the k key)
- Long-term processor averages view (pressing the l key)
- Large page analysis (pressing the L key)
- Memory and paging statistics (pressing the m key)
- Network interface view (pressing the n key)
- NFS panel (pressing the N key)
- Paging space (pressing the P key)
- age Statistics (pressing the M key)
- Disk busy map (pressing the o key)
- Disk groups (pressing the g key)
- Disk statistics (pressing the D key)
- Disk statistics with graph (pressing the d key)
- IBM ESS vpath statistics view (pressing the e key)
- FC adapter statistics (pressing the ^ key)
- JFS view (pressing the j key)
- Kernel statistics (pressing the k key)
- Long-term processor averages view (pressing the l key)
- Large page analysis (pressing the L key)
- Memory and paging statistics (pressing the m key)
- Network interface view (pressing the n key)
- NFS panel (pressing the N key)
- Paging space (pressing the P key)
- Process view (pressing the t and u keys)
- Processor usage small view (pressing the c key)
- Processor usage large view (pressing the C key)
- SEA statistics (pressing the O key)
- Shared-processor LPAR view (pressing the p key)
- System resource view (pressing the r key)
- Thread level statistics (pressing the i key)
- Verbose checks OK/Warn/Danger view (pressing the v key)
- Volume group statistics (pressing the V key)
- WLM view (pressing the W key)

For more information, see the **nmon** command, found at:

https://www.ibm.com/docs/en/aix/7.3?topic=n-nmon-command

In the recording mode, the command generates the **nmon** files. You can view these files directly by opening them or with post-processing tools such as **nmon analyzer**. The **nmon** tool disconnects from the shell during the recording so that the command continues running even if you log out.

If you use the same set of keys every time that you start the **nmon** command, you can place the keys in the NMON shell variable.

For more information, see the **nmon** recording tool commands, found at:

https://www.ibm.com/docs/en/aix/7.3?topic=tool-nmon-recording-commands

Another good tool is the VIOS Performance Advisor, which provides advisory reports that are related to the performance of various subsystems in the VIOS environment. To run this tool, run the **part** command.

For more information, see the part command, found at:

https://www.ibm.com/docs/en/power-sys-solutions/0008-DEA?topic=P8DEA/p8hcg/p8hcg_part.htm

The output that is 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 present in the `output.tar` file with the other supporting files.

To view the generated report, complete the following steps:

1. Transfer the `generated.tar` file to a system that has a browser and a `.tar` file extractor that is installed.
2. Extract the `.tar` file.
3. Open the `vios_advisor.xml` file that is in the extracted directory.

The system configuration advisory report consists of information that is related to the VIOS configuration, such as processor family, server model, number of cores, frequency at which the cores are running, and the VIOS version.

Here are the types of advisory reports that are generated by the VIOS Performance Advisor tool:

- System configuration advisory report
- CPU advisory report
- Memory advisory report
- Disk advisory report
- Disk adapter advisory report
- I/O activities (disk and network) advisory report

For more information, see the following resources:

- Virtual I/O Server Performance Advisor reports, found at:

  https://www.ibm.com/docs/en/power10/9105-22B?topic=advisor-virtual-io-server-performance-reports

- VIOS Performance Advisor called part, found at:

  https://www.ibm.com/support/pages/vios-performance-advisor-called-part

5.4 LPAR management best practices

LPAR management is a continuous process throughout the lifecycle of an LPAR. Activities like performance monitoring, dynamic logical partitioning (DLPAR), and LPAR profile changes are some of the most common tasks that are performed regularly. This section covers LPAR management aspects.

5.4.1 LPAR configuration management

You can manage the configuration of your LPARs by using the HMC. With the HMC, you can adjust the resources that are used by each LPAR.

For more information, see Managing logical partitions, found at:

https://www.ibm.com/docs/en/power10/9043-MRX?topic=hmc-managing-logical-partitions

5.4.2 LPAR performance management

LPAR performance management plays a critical role in running your workloads optimally.

For optimal performance of your LPAR, assign adequate resources as required by the workload while monitoring and analyzing the performance of the LPAR periodically.

For more information, see Performance considerations for logical partitions, found at:

https://www.ibm.com/docs/en/power10/9043-MRX?topic=hmc-performance

5.4.3 Operating systems monitoring

Operating systems monitoring is a continuous process that facilitates the tracking of resource utilization. With it, you can improve service availability and productivity by proactively and reactively identifying, diagnosing, and repairing slow, underperforming, or unstable components.

It is important to differentiate monitoring tools from management tools:

- *Monitoring tools* watch overall system, communication, and application performance, and they track various metrics of resources in the system.
- *Management tools* facilitate administering and tuning components in a system to improve system availability and performance, and ensuring that the system configuration is kept at a wanted state.

Various monitoring tools are available for Power servers and client LPARs. Some of the most popular monitoring tools are the following ones:

- nmon
- topas
- Monitoring tools for IBM i

nmon

nmon is a monitoring tool that is available for AIX and Linux LPARs.

nmon for AIX

The **nmon** tool displays system statistics in interactive mode and records system statistics in recording mode. It runs either in interactive or recording mode.

For more information, see the **nmon** command, found at:

https://www.ibm.com/docs/en/aix/7.3?topic=n-nmon-command

nmon for Linux

nmon for Linux is open source and available under the GNU General Public License.

For more information, see **nmon** for Linux, found at:

https://nmon.sourceforge.net/

topas

The **topas** command is included with AIX. It reports selected statistics about the activity on the local and remote system.

For more information, see the **topas** command, found at:

https://www.ibm.com/docs/en/aix/7.3?topic=t-topas-command

topas cross-partition view

The **topas** cross-partition view (or CEC view) allows displaying performance metrics for multiple VIOS and AIX LPARs that are running in the same Power server in real time. To view cross-partition statistics in **topas**, use the **-C** flag with the **topas** command in the CLI of any LPAR in the server.

The command displays real-time information from all LPARs on the same Power server, such as the processor and memory configuration of each LPAR and real CPU and memory utilization.

To use the **topas** cross-partition view, the following conditions must be met:

- All the LPARS must be running in the same hardware and using the same subnet.
- Make sure that the **xmtopas** service is active on all LPARS.
- Ensure that the following line is uncommented in the **/etc/inted.conf** file:

  ```
  xmquery dgram udp6 wait root /usr/bin/xmtopas xmtopas -p9
  ```

- Check the subsystems that run under the **inetd** services with the **lssrc -ls inetd** command.

For more information, see Viewing the Cross-Partition panel, found at:

https://www.ibm.com/docs/en/aix/7.3?topic=command-viewing-cross-partition-panel

Monitoring tools for IBM i

Many tools are available on IBM i to monitor, audit, and troubleshoot problems in different areas of the system:

- Monitoring security

  More than one technique is available for monitoring and auditing security on your system.

  In a security audit, you must review and examine the activities of a data processing system to test the adequacy and effectiveness of procedures for data security and data accuracy. The *security audit journal* is the primary source of auditing information on the system. A security auditor inside or outside your organization can use the auditing function that is provided by the system to gather information about security-related events that occur on the system.

An intrusion detection system is software that detects attempts or successful attacks on monitored resources that are part of a network or host system.

For more information about planning and implementing processes to monitor security on IBM i, see Monitoring security, found at:

https://www.ibm.com/docs/en/i/7.5?topic=security-monitoring

- Monitoring performance

  Several options are available on IBM i to help you identify and resolve performance problems.

  IBM iDoctor is a suite of dynamic tools that identify performance issues quickly on IBM i systems. Monitor your overall system health at a high level, and leverage advanced drill-down capabilities for specific issues. Some iDoctor tools are complementary, and others require a license key.

  For more information, see IBM iDoctor for IBM i, found at:

  https://www.ibm.com/products/idoctor

  Most of the tools that collect or analyze performance use either trace or sample data. *Collection Services* regularly collect sample data on various system resources. Several tools analyze or report on this sample data, and you can use these reports to get a broader view of system resource utilization and answer many common performance questions. *IBM i Job Watcher* and *IBM i Disk Watcher* also collect sample data. For more detailed performance information, several tools generate trace-level data. Often, trace-level data can provide detailed information about the behavior and resource consumption of jobs and applications on your system. *Performance Explorer* and the *Start Performance Trace* (**STRPFRTRC**) commands are two common tools for generating trace data.

  For more information, see Researching a performance problem, found at:

  https://www.ibm.com/docs/en/i/7.5?topic=performance-researching-problem

- Managing and administering IBM i

  IBM Navigator for i is a web console interface where you can perform the key tasks to administer your IBM i. The web application is part of the base IBM i operating system.

  For more information and a description of the management tasks that you can perform from the IBM Navigator for i console, see IBM Navigator for i, found at:

  https://www.ibm.com/docs/en/i/7.3?topic=system-navigator-i

5.5 Management solutions on PowerVM

This section describes solutions on PowerVM, which facilitate HA, migration, security, and workload optimization in PowerVM environments.

5.5.1 Migration solutions

The VIOS plays a vital role in the migration process by enabling I/O virtualization and serving during the LPM and the Simplified Remote Restart (SRR) operations:

- Partition mobility migrates AIX, IBM i, and Linux LPARs from one system to another system. The mobility process transfers the LPAR environment that includes the processor state, memory, attached virtual devices, and connected users without any shutdown or without disrupting the operation of that LPAR.

  LPM is described in 2.6.1, “Live Partition Mobility” on page 54.

  For more information, see Live Partition Mobility, found at:

  https://www.ibm.com/support/pages/live-partition-mobility

- When a server crashes, the partitions on that server also crash. To mitigate the impact of a server outage, you can use SRR capability on HMC, which is a PowerVM HA option for LPARs. SRR allows you to re-create the same LPAR with same attributes and virtual adapters in another managed system when a current managed system becomes unavailable.

  SRR is described in 2.7, “Simplified Remote Restart” on page 59.

  For more information, see Simplified Remote Restart via HMC or PowerVC, found at:

  https://www.ibm.com/support/pages/simplified-remote-restart-hmc-or-powervc

5.5.2 Availability solutions

Several availability solutions work on PowerVM:

- An *SSP* is a pool of SAN storage devices that can be used among VIOSs. SSP virtualizes FC SAN disks on the VIOS and represents them to Virtual I/O Clients over vSCSI. It enables many modern storage capabilities and functions without support from underlying disk subsystems.

  For more information about SSP, see 2.1.3, “Shared processors” on page 34, 2.1.5, “Multiple shared processor pools” on page 36, 3.2.4, “Shared processor pools capacity planning” on page 82, and 4.6, “Shared processor pools” on page 172, and the following resources:

- Creating Simple SSP among two VIO servers, found at:

  https://www.ibm.com/support/pages/creating-simple-ssp-among-two-vio-servers

- Shared Storage Pool (SSP) Best Practice, found at:

  https://www.ibm.com/support/pages/shared-storage-pool-ssp-best-practice

- IBM VM Recovery Manager (VMRM) for Power Systems is a high availability and disaster recovery (HADR) solution that enables VMs to be moved between systems by using LPM for planned operations or restarted on another system for unplanned outage events. VMs can be replicated and restarted at remote locations for disaster recovery (DR) operations.

  VMRM is described in 2.8, “VM Recovery Manager” on page 62.

For more information, see the following resources:

- IBM VM Recovery Manager for IBM Power Systems, found at:

  https://www.ibm.com/products/vm-recovery-manager

- Overview for IBM VM Recovery Manager DR for Power Systems, found at:

  https://www.ibm.com/docs/en/vmrmdr/1.6?topic=overview

- VM Recovery Manager HA overview, found at:

  https://www.ibm.com/docs/en/vmrmha/1.6?topic=overview

- IBM Power Virtualization Center (PowerVC) automated remote restart (ARR)

  The PowerVM SRR can be implemented in combination with PowerVC to provide ARR. ARR monitors hosts for a failure. If a host fails, PowerVC automatically remote restarts the virtual machines (VMs) from the failed host to another host within a host group.

  PowerVC ARR is described in 2.7.1, “PowerVC automated remote restart” on page 62.

  For more information, see Automated remote restart, found at:

  https://www.ibm.com/docs/bg/powervc/2.0.3?topic=restart-automated-remote

5.5.3 Security solutions

IBM PowerSC is a security and compliance solution that is optimized for virtualized environments on IBM Power servers that run AIX, IBM i, or Linux. PowerSC sits on top of the IBM Power server stack. It integrates security features that are built at different layers. You can centrally manage security and compliance on Power servers to get better support for compliance audits, including General Data Protection Regulation (GDPR).

For more information, see IBM PowerSC, found at:

https://www.ibm.com/products/powersc

5.5.4 Workload optimization solutions

IBM PowerVM is designed to serve high workloads. It includes several tools and enhancements to optimize the use of system resources.

Dynamic Platform Optimizer

Dynamic Platform Optimizer (DPO) is a hypervisor function that is initiated from the HMC CLI. DPO rearranges LPAR processors and memory placement on the system to improve the affinity between processors and memory resources of LPARs. When DPO is running, many virtualization features, such as mobility operations, are blocked. During a DPO operation, if you want to add, remove, or move dynamically physical memory to or from running LPARs, you must either wait for the DPO operation to complete or manually stop the DPO operation.

You can use the HMC to determine affinity scores for the system and LPARs by using the **lsmemopt** command. An affinity score is a measure of the processor-memory affinity on the system or for a partition. The score is a number 0 - 100, where 0 represents the worst affinity and 100 represents perfect affinity. Based on the system configuration, a score of 100 might not be attainable. A partition that has no processor and memory resources does not have an affinity score. When you run the **lsmemopt** command for such partitions, the score is displayed as `none` on the HMC CLI.

In addition to manually running DPO by using the **optmem** command, you can schedule DPO operations on the HMC.

The following conditions apply to a scheduled DPO operation:

- The current server affinity score of the managed system is less than or equal to the server affinity threshold that you provided.
- The affinity delta (which is the potential score minus the current score) of the managed system is greater than or equal to the affinity delta threshold of the server that you provided.

For more information, see Dynamic Platform Optimize, found at:

https://www.ibm.com/docs/en/power10?topic=dynamically-dynamic-platform-optimizer

Capacity on Demand

Servers can consist of a number of active and inactive resources. Active processor cores and active memory units are resources that are available for use on your server. Inactive processor cores and inactive memory units are resources that are installed in your server, but are not available for use until you activate them.

Capacity on Demand (CoD) is described in 2.9, “Capacity on Demand” on page 66.

For more information, see Capacity on Demand, found at:

https://www.ibm.com/docs/en/power10/9786-22H?topic=environment-capacity-demand

IBM Power Enterprise Pools

IBM Power Enterprise Pools (PEP) is an offering that is supported on certain modern Power servers. It delivers enhanced multisystem resource sharing and by-the-minute consumption of on-premises Power server compute resources to clients by deploying and managing a private cloud infrastructure on Power servers.

Two types of Power Enterprise Pool are available:

- PEP 1.0 by using Mobile Capacity.
- PEP 2.0 by using Shared Utility Capacity.

PEP is described in 2.10, “Power Enterprise Pools” on page 70.

For more information, see Power Enterprise Pool 1.0 and 2.0, found at:

https://www.ibm.com/docs/en/entitled-systems-support?topic=pools-power-enterprise-1020-overview

AIX Workload Partitions

The Workload Partition (WPAR) environment is different from the standard AIX operating system environment. Various aspects of the system, such as networking and resource controls, function differently in the WPAR environment.

The WPAR information describes how to install applications in a WPAR environment that uses various applications, such as Apache, IBM Db2®, and IBM WebSphere® Application Server. These examples are not intended to imply that they are the only supported versions or configurations of these applications.

You can create and configure application WPARs by using the **wparexec** command and the **chwpar** command.

When you create an application WPAR, a configuration profile is stored in the WPAR database. You can export this profile to create a specification file that contains the exact same configuration information for that WPAR. All WPARs must be created by an authorized administrator in the global environment.

Application WPARs provide an environment for isolation of applications and their resources to enable checkpoint, restart, and relocation at the application level. They have less usage on system resources than system WPARs and they do not require their own instance of system services.

For more information, see IBM Workload Partitions for AIX, found at:

https://www.ibm.com/docs/en/aix/7.3?topic=workload-partitions-aix

5.6 Management tools on Power servers

This section introduces some of the tools that are available to manage your Power server environment.

5.6.1 Performance and Capacity Monitor

The Performance and Capacity Monitor (PCM) is an HMC GUI that displays performance and capacity data for managed servers and LPARs.

The PCM displays data for a single physical server in a new browser window. The PCM allows the HMC to gather performance data so that a system administrator can monitor current performance and capacity changes in their Power servers environment over time.

For more information, see Performance and Capacity Monitoring, found at:

https://www.ibm.com/docs/en/power10/7063-CR2?topic=apis-performance-capacity-monitoring

5.6.2 The nmon analyzer

The **nmon_analyser** tool is helpful in analyzing performance data that is captured by using the **nmon** performance tool.

For more information, see **nmon_analyser**: A no-charge tool for producing AIX performance reports, found at:

https://developer.ibm.com/articles/au-nmon_analyser/

5.6.3 Microcode Discovery Service

Microcode Discovery Service (MDS) is used to determine whether the microcode on the Power server is at the latest level. MDS relies on an AIX utility that is called *Inventory Scout*. Inventory Scout is installed by default on all AIX LPARs.

For more information, see Microcode Discovery Service - Overview, found at:

https://esupport.ibm.com/customercare/mds/fetch?page=overview.html

As a best practice, obtain an MDS microcode report is to upload an Inventory Scout survey file for analysis.

For more information, see Microcode Discovery Service - MDS and Inventory Scout, found at:

https://esupport.ibm.com/customercare/mds/fetch?page=invreadme.html

5.6.4 Fix Level Recommendation Tool

The Fix Level Recommendation Tool (FLRT) provides cross-product compatibility information and fix recommendations for IBM products. Use FLRT to plan upgrades of key components or verify the current health of a system. Enter your current levels of firmware and software to receive a recommendation.

For more information, see FLRT, found at:

https://esupport.ibm.com/customercare/flrt/

FLRT Live Partition Mobility report

For LPM environments, the FLRT LPM report provides recommendations for LPM operations based on entered values for source and target systems. These recommendations might include fixes for known LPM issues.

For more information, see Live Partition Mobility Recommendations, found at:

https://esupport.ibm.com/customercare/flrt/lpm

Fix Level Recommendation Tool Vulnerability Checker

The Fix Level Recommendation Tool Vulnerability Checker (FLRTVC) script provides security and HIPER reports based on the inventory of your system. The FLRTVC script is a **ksh** script that uses FLRT security and HIPER data (a CSV file) to compare the installed file sets and interim fixes against known vulnerabilities and HIPER issues.

For more information, see FLRTVC Script and Documentation, found at:

https://esupport.ibm.com/customercare/flrt/sas?page=../jsp/flrtvc.jsp

5.7 PowerVC

IBM PowerVC is built on OpenStack to provide simplified virtualization management and cloud deployments for IBM AIX, IBM i, and Linux VMs. It can build private cloud capabilities on Power servers and improve administrator productivity. It can further integrate with cloud environments through higher-level cloud orchestrators.

PowerVC provides several benefits:

- Fast deployment to save time and IT costs with faster time to value through simple installation and configuration.
- Saves the cost of formal training and eliminates the need for specialized skills with an intuitive user interface.
- Cost savings and less demand on IT with resource pooling and placement policies.
- It uses host grouping to provide separate policy-based control for a subset of the total managed resources.
- It delivers policy-based automation for active workload balancing within a host group.

PowerVC offers different capabilities that depend on the edition that is installed and the hypervisor that you are using to manage your systems. For a comparison of major features and capabilities of PowerVC editions, see Feature support for PowerVC, found at:

https://www.ibm.com/docs/en/powervc/2.1.0?topic=powervc-feature-support

PowerVC enables you to capture and import images that you can deploy as VMs. An image consists of metadata and one or more binary images, one of which must be a bootable disk. To create a VM in PowerVC, you must deploy an image. For more information, see Working with images, found at:

https://www.ibm.com/docs/en/powervc/2.1.0?topic=working-images

For the most recent enhancements to PowerVC, see What's new, found at:

https://www.ibm.com/docs/en/powervc-cloud/2.1.0?topic=introduction-whats-new

For more information about PowerVC, see PowerVC 2.1.0, found at:

https://www.ibm.com/docs/en/powervc/2.1.0

6

Automation on IBM Power servers

Automation is an essential component of modernization and digital transformation on IBM Power servers. Automation is an important requirement of today's IT solutions that provides an approach to standardize the deployment, configuration, and management across the IT infrastructure. Modern infrastructure is becoming heterogeneous. IT administrators, developers, and QA engineers want to streamline anything they can do to save time and increase reliability, especially when it comes to the virtualized environment.

This chapter provides information about automation solutions that are available on Power servers and how they can be used to automate PowerVM-based deployments and configurations.

This chapter covers the following topics:

- Automation tools for Power servers
- Ansible automation for Power servers
- Automating IBM Power Virtualization Center with Ansible

6.1 Automation tools for Power servers

This section provides an overview of the options that are available for automation on Power. The focus is on automation options for AIX, IBM i, and Linux in PowerVM environments.

Here are some common automation tools:

- Puppet
- Chef
- Ansible
- Terraform

These tools can be used to automate jobs on AIX, IBM i, and Linux. Table 6-1 lists the tools that are available for each operating system.

Table 6-1 Available automation tools for each operating system

Automation tools	Virtual I/O Server (VIOS)	AIX	IBM i	Linux
Puppet	No	Yes	No	Yes
Chef	No	Yes	No	Yes
Ansible	Yes	Yes	Yes	Yes

Note: Terraform typically is not used to manage operating systems. Thus, Terraform is not listed in this table.

These tools work differently on each host operating system. The following sections provide a brief overview of each tool and references for more information.

6.1.1 Puppet

Puppet is an open-source software for IT automation. Puppet is a tool that helps you manage and automate the configuration of servers.

When you use Puppet, you define the state of the systems in the infrastructure that you want to manage. The state is defined by writing infrastructure code in the Puppet Domain-Specific Language (DSL) and Puppet code, which you can use with a wide array of devices and operating systems. Puppet code is declarative. You describe the state of your systems, not the steps that are needed to get there. Puppet automates the process of getting these systems into that state and keeping them there.

For more information, see the following resources:

- Introduction to Puppet, found at:

  https://puppet.com/docs/puppet/6/puppet_overview.html#puppet_overview
- What is Puppet, found at:

  https://www.ibm.com/support/pages/what-puppet

6.1.2 Chef

The Chef software is an open-source configuration management tool that you can use to create parts of an infrastructure as a service (IaaS). It works as client/server model architecture. The Chef procedural script language creates *recipes*, which are made of reusable definitions that are written in the Ruby programming language. Chef recipes that perform related functions are grouped in a single container that is called a cookbook. Cookbooks and recipes automate common infrastructure tasks. The recipe definitions describe what your infrastructure consists of, and how each part of your infrastructure is deployed, configured, and managed. Chef applies these definitions to servers to produce an automated infrastructure.

Using Chef, you can create a scalable infrastructure with a minimal configuration to maintain. Chef scripts form part of a recipe for your infrastructure configuration that is understood by multiple systems, services, and languages. By using Chef to configure your infrastructure deployment, you do not have to rewrite code if you change part of the underlying infrastructure. For more information, see the following resources:

- Chef Platform Overview, found at:

  https://docs.chef.io/platform_overview/

- Chef, found at:

  https://www.ibm.com/docs/en/integration-bus/9.0.0?topic=service-chef

- About Cookbooks, found at:

  https://docs.chef.io/cookbooks/

- *Porting Chef Cookbooks to AIX*, found at:

  https://www.ibm.com/support/pages/system/files/inline-files/IBM_AIX_Chef_porting_Kohlmeier_09NOV15.pdf

6.1.3 Ansible

Ansible is an open-source IT configuration management, deployment, and orchestration tool.

Ansible is the automation platform of choice for Power server. For this reason, 6.2, “Ansible automation for Power servers” on page 210 describes the Ansible Automation Platform in more detail and provides references to more resources for further reading.

For more information, see *ANSIBLE IN DEPTH*, found at:

https://www.ansible.com/hubfs/pdfs/Ansible-InDepth-WhitePaper.pdf

6.1.4 Terraform

Terraform is an open-source “Infrastructure as Code” tool that was created by HashiCorp. Terraform is a declarative coding tool that enables developers to use a high-level configuration language called HashiCorp Configuration Language (HCL). With this tool, developers describe the wanted “end-state” cloud or on-premises infrastructure for running an application. Then, HCL generates a plan for reaching that end state and runs the plan to provision the infrastructure.

For more information, see Terraform, found at:

https://www.ibm.com/in-en/cloud/learn/terraform

6.2 Ansible automation for Power servers

Ansible is an IT automation tool. Ansible is used to install, configure, and deploy systems and software, and orchestrate common IT tasks, such as continuous deployments or zero downtime rolling updates.

Simplicity, security, and reliability are Ansible's main objectives.

Ansible uses OpenSSH for transport, the open-source connectivity tool for remote login with the Secure Shell (SSH) protocol. Other transports and pull modes can be used as alternatives. The use of OpenSSH reduces security exposures.

Ansible uses a programming language that can be read and interpreted by humans, even by users that are not familiar with the programming language.

Ansible facilitates an agentless management of machines, which eliminates issues that are related to endpoint daemon management. Instead, Ansible runs with a *push* model.

Ansible manages remote machines over SSH. Because Ansible is an agentless tool, it does not consume any resources on the endpoints when not in use. Also, it does not need a dedicated Ansible administrator on the endpoints. When required, Ansible can use **`sudo`**, **`su`**, and other privilege escalation methods.

Automating IT operations with Ansible delivers many benefits to IT managers and CIOs. Organizations benefit from improvements in efficiency, visibility, and simplicity. The resulting high level of repeatable automated IT processes increases productivity (admins can get more done) and reduces risk (less operator error). Ansible automated management also spans multiple configurable endpoints, such as servers, VMs, network switches, storage devices, and platforms, including Power servers, mainframe, and x86.

Ansible brings together administrators and developers to collaborate on building IT automation solutions that work for them. Ansible has a vibrant community that is constantly innovating and delivering new capabilities to extend its reach within the data center.

For more information, see the following resources:

- Ansible Documentation, found at:

  `https://docs.ansible.com/ansible/latest/index.html`
- IBM Power Systems and the Red Hat Ansible Automation Platform, found at:

  `https://www.ansible.com/integrations/infrastructure/ibm-power-systems`
- Ansible for IBM Power Systems (video), found at:

  `https://mediacenter.ibm.com/media/t/1_ijxrlhjh`
- Community Information & Contributing, found at:

  `https://docs.ansible.com/ansible/2.3/community.html#community-information-contributing`

6.2.1 Ansible Content Collections

Ansible provides content in a distribution format that is called Ansible Content Collections, or collections, which represent a new standard of distributing, maintaining, and consuming automation. By combining multiple types of Ansible content (playbooks, roles, modules, and plug-ins), flexibility and scalability are improved.

These collections are available at using Ansible Galaxy (community version) and Ansible Automation Hub.

For more information, see Getting Started With Ansible Content Collections, found at:

https://www.ansible.com/blog/getting-started-with-ansible-collections

IBM Collections for Ansible

IBM Collections for Power servers provide collection for AIX, VIOS, Hardware Management Console (HMC), and IBM i. They are available at Ansible Galaxy.

Figure 6-1 shows the available Power servers collections at Ansible Galaxy.

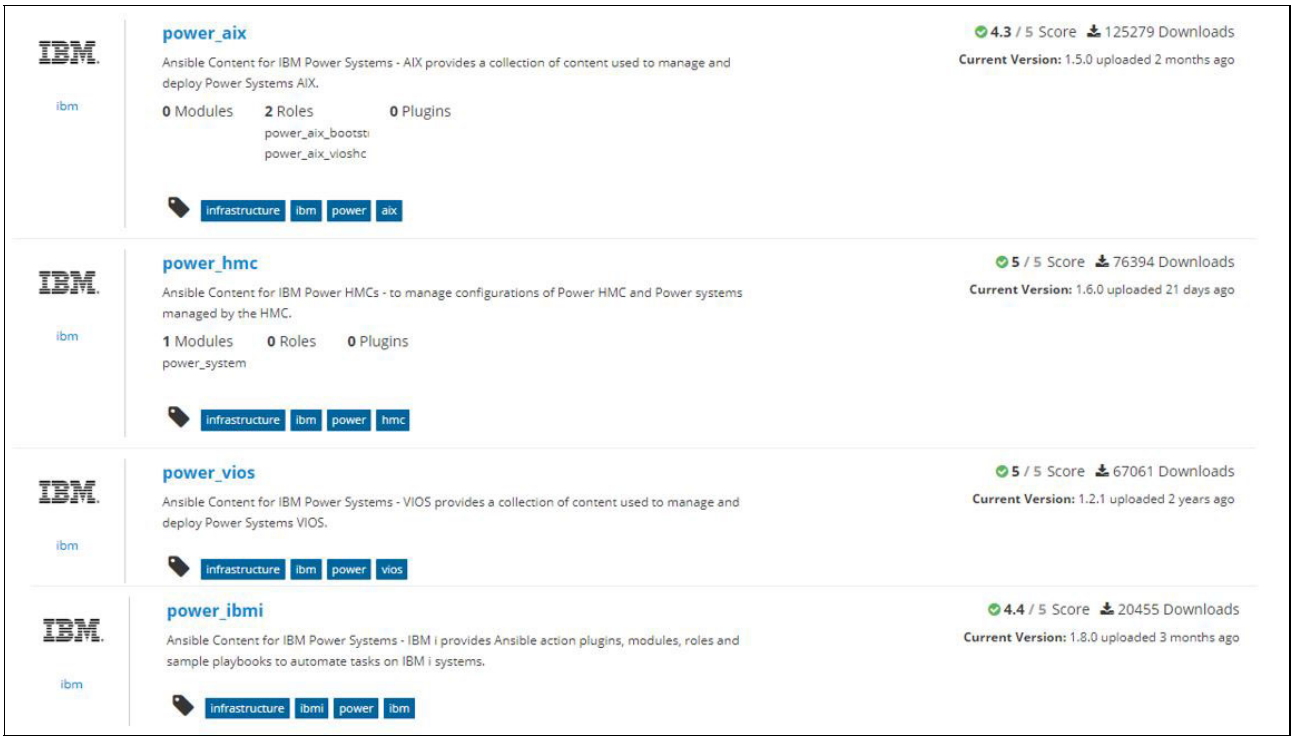


Figure 6-1 Ansible Galaxy Power servers collections

The following Ansible collections are available for Power servers:

- IBM Power Systems AIX Collection (`power_aix`)

  The IBM Power Systems AIX collection provides modules that can be used to manage configurations and deployments of AIX on Power servers and AIX logical partitions (LPARs). The collection content helps to include workloads on Power platforms as part of an enterprise automation strategy through the Ansible ecosystem.

  For more information, see IBM Power Systems AIX Collection, found at:

  https://galaxy.ansible.com/ibm/power_aix

- IBM Power Systems HMC Collection (`power_hmc`)

  The IBM Power Systems HMC collection provides modules that can be used to manage configurations and deployments of HMC on Power servers and Power servers that are managed by the HMC. The collection content helps to include workloads on Power platforms as part of an enterprise automation strategy through the Ansible ecosystem.

For more information, see IBM Power Systems HMC Collection, found at:

https://galaxy.ansible.com/ibm/power_hmc

- IBM Power Systems VIOS Collection (`power_vios`)

  The IBM Power Systems VIOS collection provides modules that can be used to manage configurations and deployments of VIOS on Power servers. The collection content helps to include workloads on Power platforms as part of an enterprise automation strategy through the Ansible ecosystem.

  For more information, see IBM Power Systems VIOS Collection, found at:

  https://galaxy.ansible.com/ibm/power_vios

- Ansible Content for IBM Power Systems - IBM i (`power_ibmi`)

  The Ansible Content for IBM Power Systems - IBM i provides modules, action plug-ins, roles, and sample playbooks to automate tasks on IBM i, such as command runs, system and application configuration, work management, fix management, and application deployment.

  For more information, see Ansible Content for IBM Power Systems - IBM i, found at:

  https://galaxy.ansible.com/ibm/power_ibmi

For more information, see the following resources:

- Ansible Galaxy – Open-source repository of Power modules, found at:

  https://galaxy.ansible.com/search?keywords=ibm%20and%20power&order_by=-relevance

- Automate AIX and IBM i Admin Tasks with Ansible Content, found at:

  https://www.ansible.com/resources/webinars-training/automate-aix-and-ibm-i-admin-tasks-with-ansible-content-webinar

- Ansible for AIX demo, found at:

  https://mediacenter.ibm.com/media/t/1_2furx7gl

- Ansible for IBM i demo, found at:

  https://mediacenter.ibm.com/media/t/1_fdz7x3vi

6.2.2 Ansible Automation Platform 2 for IBM Power Systems

The Ansible Automation Platform is a foundation for building and operating automation across an organization. The platform includes all the tools that are required to implement enterprise-wide automation. Ansible-managed endpoints can include virtual machines (VMs) on Power servers that run IBM AIX, IBM i, or Linux.

The Ansible Automation Platform provides enterprise subscriptions and support for Ansible deployments, including certified endpoint module collections for AIX, IBM i, Linux, VIOS, and HMC.

The Ansible Automation Platform for IBM Power Systems makes it possible for users across an organization to create, test, and manage automation content through a powerful and agentless framework. It is a more secure, stable, and flexible foundation for deploying end-to-end automation solutions, from IT processes to hybrid cloud to the edge.

Here are some of the key features:

- IT managers and architects can more easily expand automation across the enterprise. Automation policy and governance are managed with the automation services catalog. They can get real-time reporting across the entire stack.
- Execution environments deliver a consistent container-like experience for building and scaling automation. New tools are included to help build and manage them. Ansible Content Collections offer prebuilt automation content from more than 100 certified partners, with solutions that are available for many use cases.
- Administrators and operators have powerful tools in the automation controller and automation hub to manage and share automation projects more efficiently, with a common language and broadly accessible mix of command-line interfaces (CLIs), GUIs, and text-based user interfaces (TUIs) across endpoints.

Red Hat Ansible Certified Content for IBM Power Systems is available as part of the enterprise automation strategy.

Figure 6-2 provides an overview of Ansible Automation Platform2 for IBM Power Systems.

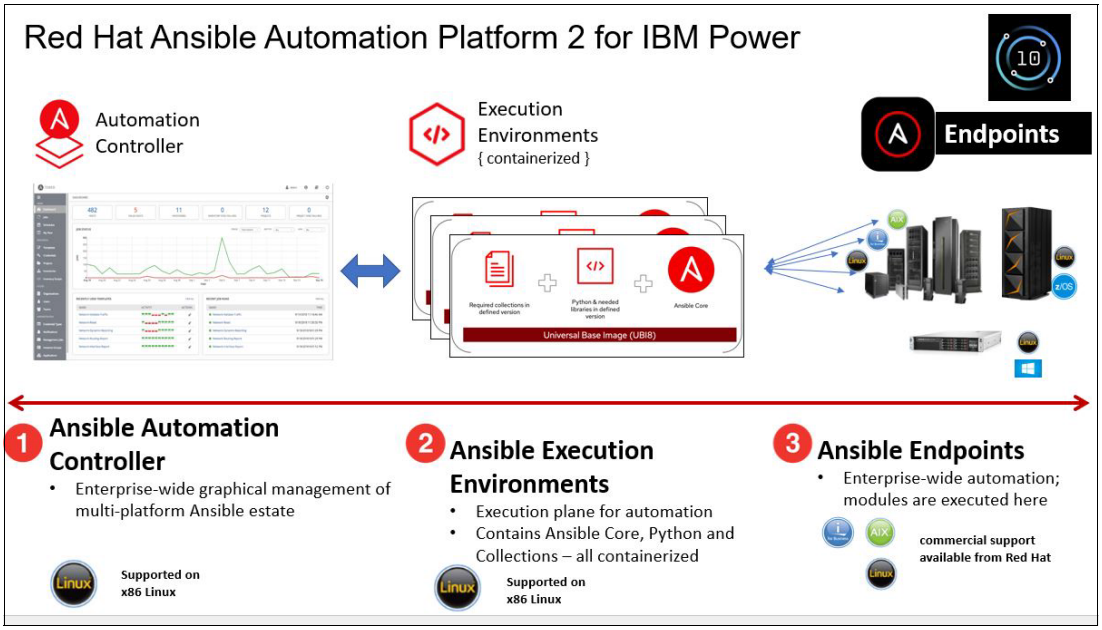


Figure 6-2 Ansible Automation Platform for IBM Power Systems

For more information, see the following resources:

- IBM enables Red Hat Ansible Automation Platform for IBM Power, found at:
  https://www.ibm.com/downloads/cas/US-ENUS222-044-CA
- CERTIFIED INTEGRATION: Ansible and IBM Power Systems, found at:
  https://www.ansible.com/integrations/infrastructure/ibm-power-systems
- Red Hat Ansible Automation Platform trial, found at:
  https://www.redhat.com/en/technologies/management/ansible/trial

6.3 Automating IBM Power Virtualization Center with Ansible

Automation is a foundation for digital transformation. Red Hat Ansible is a powerful tool that plays an important role in a digital transformation journey.

IBM Power Virtualization Center (PowerVC) is the strategic enterprise virtualization and cloud management solution for Power servers. It provides several benefits, such as simplified virtualization management and operations, rapid provisioning, and upward integration with other technologies, which include Ansible. PowerVC is built on OpenStack technology. Because of its OpenStack foundation, the OpenStack modules that are included with Ansible are all that you need to get started.

Note: These Ansible modules are available through the community only, that is, no enterprise support is provided.

You can use the Ansible `uri` module to directly call PowerVC APIs that might not be exposed through the OpenStack modules.

The tutorial Automating PowerVC using Ansible illustrates a practical application of automating IBM PowerVC with Ansible to provision a VM. For implementation details, see the tutorial, found at:

https://developer.ibm.com/tutorials/automating-powervc-using-ansible/

For more information, see the following resources:

- Automating PowerVC snapshots using Ansible and REST API, found at:

  https://community.ibm.com/community/user/power/blogs/nicolae-chirea/2022/05/27/powervc-snapshots-ansible-rest

- Ansible Role for PowerVC AIX VM Deployment - GitHub, found at:

  https://github.com/lg4U/Ansible_PowerVC_deployVM

Abbreviations and acronyms

AME	Active Memory Expansion
AMM	Active Memory Mirroring
ARP	Address Resolution Protocol
ARR	automated remote restart
ASMI	Advanced System Management Interface
BMC	Baseboard Management Controller
BOOTP	Boot Protocol
CAA	Cluster Aware AIX
CAM	Cloud Automation Manager
CCEVS	Common Criteria Evaluation and Validation Scheme
CIS	Center for Internet Security
CLI	command-line interface
CoD	Capacity on Demand
CUoD	Capacity Upgrade on Demand
DGD	Dead Gateway Detection
DHCP	Dynamic Host Configuration Protocol
DLPAR	dynamic logical partitioning
DMZ	demilitarized zone
DoD	Department of Defense
DPO	Dynamic Platform Optimizer
DR	disaster recovery
DSL	Domain-Specific Language
DSS	Data Security Standard
eBMC	enterprise Baseboard Management Controller
EPC	Entitled Pool Capacity
FC	Fibre Channel
FCP	Fibre Channel Protocol
FHE	Fully Homomorphic Encryption
FLRT	Fix Level Recommendation Tool
FLRTVC	Fix Level Recommendation Tool Vulnerability Checker
FQDN	fully qualified domain name
FSP	Flexible Service Processor
FTP	File Transfer Protocol
GDPR	General Data Protection Regulation
GDR	Geographically Dispersed Resiliency
HA	high availability
HADR	high availability and disaster recovery
HBA	host bus adapter
HCA	Host Channel Adapter
HCL	HashiCorp Configuration Language
HEA	Host Ethernet Adapter
HIPAA	Health Insurance Portability and Accountability Act
HIPER	High Impact PERvasive
HMC	Hardware Management Console
HNV	Hybrid Network Virtualization
HPT	hardware page table
IaaS	infrastructure as a service
IBM	International Business Machines Corporation
IBM CMC	IBM Cloud Management Console
IBM ESS	IBM Entitled Systems Support
IPAT	IP Address Takeover
IPL	initial program load
ISA	Instruction Set Architecture
ISV	Independent software vendor
IVM	Integrated Virtualization Manager
LA	link aggregation
LaMa	Landscape Management
LIC	Licensed Internal Code
LMB	logical memory block
LP	logical port
LPAR	logical partition
LPM	Live Partition Mobility
LSO	large send offload
LU	logical unit
LUN	logical unit number
LVM	logical volume mirroring
MCP	Management Control Point
MDS	Microcode Discovery Service
MFA	Multifactor Authentication
MPC	Maximum Pool Capacity
MPIO	multipath input/output
MSP	mover service partition
MSPP	multiple shared processor pools
MTU	maximum transmission unit

NAS	network-attached storage
NAT	Network Address Translation
NDP	Neighbor Discovery Protocol
NFV	network function virtualization
NIB	Network Interface Backup
NIC	network interface card
NIM	Network Installation Manager
NPIV	N_Port ID Virtualization
NTP	Network Time Protocol
OLTP	online transaction processing
OSPF	Open Shortest Path First
PCI	Payment Card Industry
PCIe	Peripheral Component Interconnect Express
PCM	Performance and Capacity Monitor
PEP	IBM Power Enterprise Pools
PHYP	PowerVM hypervisor
PKS	Platform keystore
PowerVC	IBM Power Virtualization Center
PowerVS	IBM Power Virtual Server
PRS	Platform Resource Scheduler
QoS	quality of service
RAS	reliability, availability, and serviceability
RBAC	role-based access control
RCA	root cause analysis
REST	Representational State Transfer
RFP	request for proposal
RMC	Resource Monitoring and Control
RPC	Reserved Pool Capacity
RSCT	Reliable Scalable Cluster Technology
RSS	receive side scaling
RSTP	Rapid Spanning-Tree
SAN	storage area network
SDN	software-defined networking
SEA	Shared Ethernet Adapter
SME	subject matter expert
SMS	System Management Services
SMT	simultaneous multithreading
SOW	statement of work
SP	Service Pack
SPOF	single point of failure
SPP	shared processor pool
SPT	System Planning Tool
SR-IOV	single-root I/O virtualization
SRR	Simplified Remote Restart
SSH	Secure Shell
SSP	shared storage pool
ST	single-threaded
STIG	Security Technical Implementation Guide
SUE	Special Uncorrectable Error
SVFC	server virtual Fibre Channel
TCE	Translation Control Entry
TME	Transparent Memory Encryption
TPM	Trusted Platform Module
TUI	text-based user interface
UDID	unique device ID
USB	Universal Serial Bus
VEA	Virtual Ethernet Adapter
VEB	Virtual Ethernet Bridge
VEPA	Virtual Ethernet Port Aggregator
VF	virtual function
VFC	virtual Fibre Channel
vHMC	virtual Hardware Management Console
VI	virtual intermediary
VIOS	Virtual I/O Server
VIPA	virtual IP address
VLAN	virtual local area network
VM	virtual machine
VMI	Virtualization Management Interface
VMRM	VM Recovery Manager
vNIC	virtual Network Interface Controller
vSCSI	virtual SCSI
VTD	virtual target device
vTPM	Virtual Trusted Platform Module
WPAR	Workload Partition
WWPN	worldwide port name

Related publications

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

IBM Redbooks

The following IBM Redbooks publications provide additional information about the topics in this document. Some publications that are referenced in this list might be available in softcopy only.

- *IBM Power E1050: Technical Overview and Introduction*, REDP-5684
- *IBM Power E1080 Technical Overview and Introduction*, REDP-5649
- *IBM Power S1014, S1022s, S1022, and S1024 Technical Overview and Introduction*, REDP-5675
- *IBM Power Systems Private Cloud with Shared Utility Capacity: Featuring Power Enterprise Pools 2.0*, SG24-8478
- *IBM PowerVM Virtualization Introduction and Configuration*, SG24-7940
- *IBM PowerVM Virtualization Managing and Monitoring*, SG24-7590
- *Implementing IBM VM Recovery Manager for IBM Power Systems*, SG24-8426

You can search for, view, download, or order these documents and other Redbooks, Redpapers, web docs, drafts, and additional materials, at the following website:

ibm.com/redbooks

Online resources

These websites are also relevant as further information sources:

- AIX I: IBM PowerVM Logical Partition Management (course AN11DG):

  https://www.ibm.com/training/course/AN11DG
- IBM Power10 processor-based servers documentation:

  https://www.ibm.com/docs/en/power10
- IBM Power Community:

  https://community.ibm.com/community/user/power/home
- IBM Servers Twitter:

  https://twitter.com/ibmservers?lang=en
- Implementing PowerVM Live Partition Mobility (course AN33G):

  https://www.ibm.com/training/course/AN33G

- Power Systems for AIX - PowerVM I Implementing Virtualization (course AN30G):

  https://www.ibm.com/training/course/AN30G
- Power Systems for AIX - Virtualization II: Advanced PowerVM and Performance (course AN31G):

  https://www.ibm.com/training/course/AN31G

Help from IBM

IBM Support and downloads

ibm.com/support

IBM Global Services

ibm.com/services

Redbooks

Introduction to IBM PowerVM

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

SG24-8535-00

ISBN 0738461024

Printed in U.S.A.