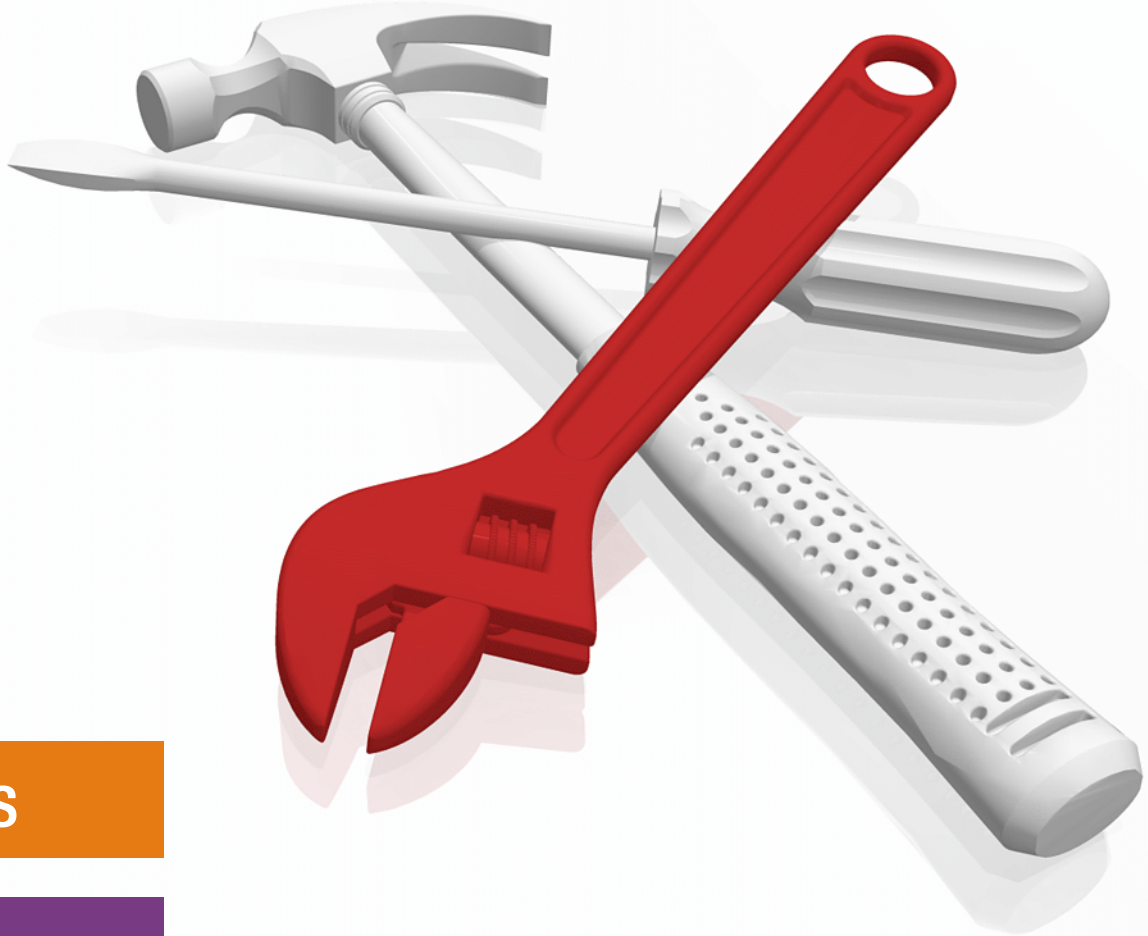
Redbooks
ibm.com/redbooks

IBM Power Systems Virtualization Operation Management for SAP Applications

Dino Quintero
Enrico Joedecke
Katharina Probst
Andreas Schauberer



Analytics

Power Systems

IBM

Redpaper

IBM Redbooks

IBM Power Systems Virtualization Operation Management for SAP Applications

March 2020

REDP-5579-00

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

First Edition (March 2020)

This edition applies to the following products:

- Red Hat Enterprise Linux 7.6
- Red Hat Virtualization 4.2
- SUSE Linux SLES 12 SP3
- HMC V9 R1.920.0
- Novalink 1.0.0.10
- ipmitool V1.8.18

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®
IBM®
Micro-Partitioning®
POWER®
POWER7®
POWER8®
POWER9™
PowerVM®
Redbooks®
Redbooks (logo)  ®

The following terms are trademarks of other companies:

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

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.

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

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

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

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

Preface

Businesses are using IBM® Power Systems servers and Linux to consolidate multiple SAP workloads onto fewer systems, increasing infrastructure utilization; reliability, availability, and serviceability (RAS); and scalability, and reducing cost.

This IBM Redpaper publication describes key hardware and software components of an SAP solution stack. Furthermore, this book addresses non-functional items like RAS, security, and issue handling. Practical help for planning, implementation, configuration, installation, and monitoring of a solution stack are provided.

This publication addresses topics for sellers, IT architects, IT specialists, and anyone who wants to implement and manage SAP workloads on IBM Power Systems servers. Moreover, this guide provides documentation to transfer how-to skills to the technical teams, and it provides solution guidance to the sales team. This publication complements documentation that is available at IBM Knowledge Center, and it aligns with educational materials that are provided by IBM Systems.

Authors

This paper was produced in close collaboration by the IBM SAP International Competence Center (ISICC) in Walldorf, SAP Headquarters in Germany and IBM Redbooks®.

Dino Quintero is an IT Management Consultant and an IBM Level 3 Senior Certified IT Specialist with IBM Redbooks in Poughkeepsie, New York. He has 24 years of experience with IBM Power Systems technologies and solutions. Dino shares his technical computing passion and expertise by leading teams developing technical content in the areas of enterprise continuous availability, enterprise systems management, high-performance computing (HPC), cloud computing, artificial intelligence (AI) (including machine learning and deep learning), and cognitive solutions. He is a Certified Open Group Distinguished IT Specialist. Dino holds a Master of Computing Information Systems degree and a bachelor of science degree in computer science from Marist College.

Enrico Joedecke is a Senior Software Engineer working for the IBM Systems Lab in Germany. He has 20 years of experience with the IBM POWER® platform, IBM AIX®, Linux, which are provided for SAP applications. In earlier years, he worked in different software engineering roles for IBM high availability (HA) software for SAP applications. Currently, he is a member of the IBM HANA On POWER Development Team in Sankt-Leon-Rot. Enrico holds a German degree of “Diplom Informatiker” (comparable to Master of Computer Science) from the University of Leipzig.

Katharina Probst leads the IBM Development team inside the SAP Headquarters, which is responsible for the data center readiness and ecosystem enablement of SAP HANA on Power Systems. She has 15 years of experience with SAP on Power Systems in the areas of storage, business continuity, disaster recovery (DR), AIX, and Linux.

Andreas Schauberer is a Senior Software Engineer working for the IBM Systems Lab in Germany. He has 15 years of experience with the IBM POWER platform, IBM AIX, Linux, which are provided for SAP applications. In earlier years, he worked in different software engineering roles for IBM high availability (HA) software for SAP applications. In his current role, he leads the IBM development team that is responsible for SAP NetWeaver and S/4HANA Foundation on IBM POWER Linux platform. Andreas holds a German degree of “Diplom Informatiker (FH)” from Fachhochschule Giessen.

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

Wade Wallace
IBM Redbooks, Austin Center

Markus Butsch, Walter Orb, Tanja Scheller
IBM Germany

Sahitya Kumar
IBM India

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 papers to be as helpful as possible. Send us your comments about this paper 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 Facebook:

  http://www.facebook.com/IBMRedbooks

- Follow us on Twitter:

  http://twitter.com/ibmredbooks

- Look for us on LinkedIn:

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

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

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

- Stay current on recent Redbooks publications with RSS Feeds:

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

1

Introduction

This chapter provides a brief summary of the contents of the publication.

This chapter contains the following topic:

- Preface

1.1 Preface

Businesses are using IBM Power Systems servers and Linux to consolidate multiple SAP workloads onto fewer systems, increasing infrastructure utilization; reliability, availability, and serviceability (RAS); and scalability, and reducing cost.

This publication describes key hardware and software components of an SAP solution stack in Chapter 2, “Server virtualization” on page 3 and Chapter 4, “Kernel-based virtual machine and Red Hat Virtualization management and operations” on page 53. Furthermore, this book addresses non-functional items such as RAS and security in Chapter 3, “IBM PowerVM management and operations” on page 19.

Due to the complexity of the topic, this publication cannot cover all aspects of it in full detail. Instead, it provides an introduction to the topic and refers to other publications for more information.

2

Server virtualization

This chapter describes the IBM Power Systems servers virtualization features.

This chapter contains the following topics:

- Introduction
- Server and hypervisor options
- Hypervisors

2.1 Introduction

Server virtualization is the process of using software (a so-called *hypervisor*) on a physical server to create multiple partitions or *virtual instances*, each of which can run independently, unlike a single dedicated server, where the entire machine has only one instance of an operating system (OS). On a virtual server, the same machine can be used to run multiple server instances, each with independent OS configurations and separate CPU core, memory, and disk configurations.

2.2 Server and hypervisor options

On IBM POWER8® and POWER9™ processor-based systems, the customer must choose between two virtualization technologies and select the hardware model during the ordering process. The two virtualization technologies are IBM PowerVM® hypervisor and the open source kernel-based virtual machine (KVM). The virtualization technology cannot be changed after the hardware is sent to the customer.

2.2.1 Power Systems models that support PowerVM versus KVM

The focus of the IBM PowerVM hypervisor is high performance, RAS, and scalability for the OSs IBM AIX, IBM i, and Linux. KVM is focused on providing fast virtualization for Linux guests on small- and medium-sized systems. IBM PowerKVM was the first product that supported KVM on Power Systems, and it had support for SAP solutions. However, this product is no longer supported by IBM (see IBM United States Withdrawal Announcement 916-17). At the time of writing, Red Hat Virtualization (RHV) supports SAP solutions that use KVM technology on Power Systems. Support for SAP NetWeaver on Red Hat KVM / RHV is documented in SAP Note 14009111.

2.2.2 Overview of POWER8 and POWER9 processor-based hardware models

This section provides an overview of IBM POWER9 processor-based server models, as shown in Table 2-1.

Table 2-1 POWER9 processor-based server models

Power Systems server	Model number	Processor	Max cores	Max RAM	IBM PowerVM	KVM / RHV	IBM Redbooks publication
E980	9080-M9S	POWER9	192	64 TB	X		*IBM Power System E980: Technical Overview and Introduction*, REDP-5510
E950	9040-MR9	POWER9	48	16 TB	X		*IBM Power System E950: Technical Overview and Introduction*, REDP-5509
S924	9009-42A	POWER9	24	4 TB	X		*IBM Power Systems S922, S914, and S924 Technical Overview and Introduction*, REDP-5497

H924	9223-42H	POWER9	24	4 TB	X		*IBM Power Systems H922 and H924 Technical Overview and Introduction*, REDP-5498
S922	9009-22A	POWER9	20	4 TB	X		*IBM Power Systems S922, S914, and S924 Technical Overview and Introduction*, REDP-5497
H922	9223-22H	POWER9	20	4 TB	X		*IBM Power Systems H922 and H924 Technical Overview and Introduction*, REDP-5498
S914	9009-41A	POWER9	8	1 TB	X		*IBM Power Systems S922, S914, and S924 Technical Overview and Introduction*, REDP-5497
L922	9008-22L	POWER9	24	4 TB	X		*IBM Power System L922 Technical Overview and Introduction*, REDP-5496
LC921	9006-12P	POWER9	40	2 TB		X	*IBM Power Systems LC921 and LC922: Technical Overview and Introduction*, REDP-5495
LC922	9006-22P	POWER9	44	2 TB		X	*IBM Power Systems LC921 and LC922: Technical Overview and Introduction*, REDP-5495

IBM POWER8 processor-based server models are shown in Table 2-2.

Table 2-2 POWER8 server models

Power Systems server	**Model number**	**Processor**	**Max cores**	**Max RAM**	**IBM PowerVM**	**KVM / RHV**	**IBM Redbooks publication**
E880C	9080-MHE	POWER8	192	32 TB	X		*IBM Power Systems E870C and E880C Technical Overview and Introduction*, REDP-5413
E870C	9080-MME	POWER8	64	16 TB	X		*IBM Power Systems E870C and E880C Technical Overview and Introduction*, REDP-5413
E880	9119-MHE	POWER8	192	32 TB	X		*IBM Power Systems E870 and E880 Technical Overview and Introduction*, REDP-5137
E870	9119-MME	POWER8	80	16 TB	X		*IBM Power Systems E870 and E880 Technical Overview and Introduction*, REDP-5137

E850C	8408-44E	POWER8	48	4 TB	X		*IBM Power System E850C Technical Overview and Introduction*, REDP-5412
E850	8408-E8E	POWER8	48	4 TB	X		*IBM Power System E850 Technical Overview and Introduction*, REDP-5222
S824	8286-42A	POWER8	24	2 TB	X		*IBM Power Systems S814 and S824 Technical Overview and Introduction*, REDP-5097
S822	8284-22A	POWER8	20	1 TB	X		*IBM Power System S822 Technical Overview and Introduction*, REDP-5102
S814	8286-41A	POWER8	8	1 TB	X		*IBM Power Systems S814 and S824 Technical Overview and Introduction*, REDP-5097
S824L	8247-42L	POWER8	24	2 TB	X	X	*IBM Power System S824L Technical Overview and Introduction*, REDP-5139
S822L	8247-22L	POWER8	24	1 TB	X	X	*IBM Power Systems S812L and S822L Technical Overview and Introduction*, REDP-5098
S812L	8247-21L	POWER8	12	512 GB	X	X	*IBM Power Systems S812L and S822L Technical Overview and Introduction*, REDP-5098
S822LC	8335-GCA 8335-GTA	POWER8	20	1 TB		X	*IBM Power System S822LC Technical Overview and Introduction*, REDP-5283
S822LC	8001-22C	POWER8	22	512 GB		X	*IBM XIV Security with Data-at-Rest Encryption*, REDP-5047
S812LC	8348-21C 8001-12C	POWER8	10	1 TB		X	*IBM Power System S812LC Technical Overview and Introduction*, REDP-5284

2.2.3 Comparison of PowerVM and KVM / RHV

Example 2-1 compares IBM PowerVM and KVM / RHV features.

Example 2-1 PowerVM and KVM / RHV features

Feature	IBM PowerVM	KVM/RHV
Adding devices to the guest	Dynamic logical partition (DLPAR)	Hot plug
Different editions	Yes (Standard and Enterprise)	Yes, Red Hat Enterprise Linux (RHEL) (<= 4 guests) and RHV (1 - *n* guests)
DLPAR	Yes	Yes
Guests running in Big and Little Endian	Yes	Yes
License	Proprietary	Open source
Live Partition Mobility (LPM)	Yes	Yes
Memory compression	Yes (IBM Active Memory Expansion)	No (Zswap can be installed manually.)
Memory page sharing	Yes (Described as Active Memory Data Deduplication.)	Yes (Described as Kernel SamePage Merging (KSM).)
IBM Micro-Partitioning® / Shared processors	Yes	Yes
Shared processor pools (SPPs)	Yes	No
Guaranteed minimum entitlement	Yes	No
Hard capping of virtual machines (VMs)	Yes	No
Capacity on Demand (CoD)	Yes	No
N-Port ID Virtualization (NPIV)	Yes	Yes
PCI pass-through	Yes	Yes
Shared storage pools	Yes	Yes
Sparse disk storage	Yes (thin provisioning)	Yes (qcow2 image)
Supported machines	All non-LC IBM Power Systems	All LC model Power Systems servers and some L model Power Systems servers
Supported OSs in the guest	IBM AIX, IBM i, and Linux	Linux

2.3 Hypervisors

This section describes the hypervisors that are available in Power Systems servers.

2.3.1 Introducing IBM PowerVM

PowerVM is the virtualization technology that is offered by Power Systems. PowerVM server virtualization can consolidate multiple workloads onto fewer systems, increase server utilization, and reduce cost. PowerVM provides a secure and scalable server virtualization environment for AIX, IBM i, and Linux OSs that are built on the advanced Remote Access Service features and leading performance of the Power platform. Multiple different PowerVM features like LPM, Live Kernel Update (LKU), and single root input/output virtualization (SR-IOV) can be deployed by customers to meet their enterprise requirements. PowerVM easily handles all virtualization aspects that are needed in a data center.

The Hardware Management Console (HMC) can be used to easily create and administer logical partitions (LPARs) on Power Systems by using either a GUI or a simple command-line interface (CLI). The HMC also provides tools to monitor the performance and change resource allocations at any time. Advanced features like DLPAR configuration and Active Memory Expansion (AME) can be used to adjust the CPU and memory usage of an LPAR. Dynamic Processor Optimizer (DPO) can be used to manage the LPAR CPU and memory placements to improve their affinity. All of these functions can be managed easily through the HMC, and are described in more detail in the following sections.

A Virtual I/O Server (VIOS) is a special partition that is provided by PowerVM. This LPAR facilitates sharing of physical I/O resources among client LPARs within the server. The VIOS provides storage virtualization by using virtual SCSI and virtual Fibre Channel. Network virtualization is provided by using Shared Ethernet Adapter (SEA), SR-IOV, virtual Network Interface Controller (vNIC), or SR-IOV Virtual Functions (SR-IOV-VF). The VIOS is also used to provide the Active Memory Sharing capability to client LPARs within the system, and suspend/resume and remote restart features to AIX, IBM i, and Linux. It can be configured to provide failover and disaster recovery (DR) functions to LPARS. Physical storage and network adapters are typically assigned directly to the VIOS to make them available to all LPARs on the server, which is a prerequisite for LPM, remote/restart, and other DR functions that are used in PowerVM. A VIOS also reduces the number of physical adapters that are required for Power Systems servers because a single adapter can be used by multiple LPARs, which increases the utilization of an adapter and reduces data center infrastructure complexity.

The following sections describe important PowerVM features and how they relate to SAP and the Linux OS.

LPAR virtualization options

Logical partitioning is a feature that is provided by PowerVM, which is used to separate the resources of a physical system into subsets that are called LPARs. Processors, memory, and input/output devices can be individually assigned to LPARs, or LPARs can share the resources from a common pool. Each LPAR runs its own version of an OS because LPARs run as independent logical servers with the resources that are allocated to them.

PowerVM allows the creation of different types of LPARs that have unique characteristics. These types are dedicated LPARs or shared processor LPARs (SPLPARs) with or without Micro-Partitioning.

Dedicated LPARs

Dedicated LPARs allocate their CPU resources for exclusive use, One or more physical processors are fully owned by this LPAR. So, the number of dedicated LPARs on a Power Systems server is limited by the number of available physical processors. The processor capacity of dedicated LPARs must be high enough to fulfill the requirements of the expected peak load. If a peak load is higher than the configured CPU resources, other processing capacity in the server might remain unused.

CPUs in a dedicated LPAR are reserved for only this LPAR, but the LPAR can be configured so that it donates unused processor cycles to a configured SPP. You can set this configuration within the partition profile, as shown in Figure 2-1.

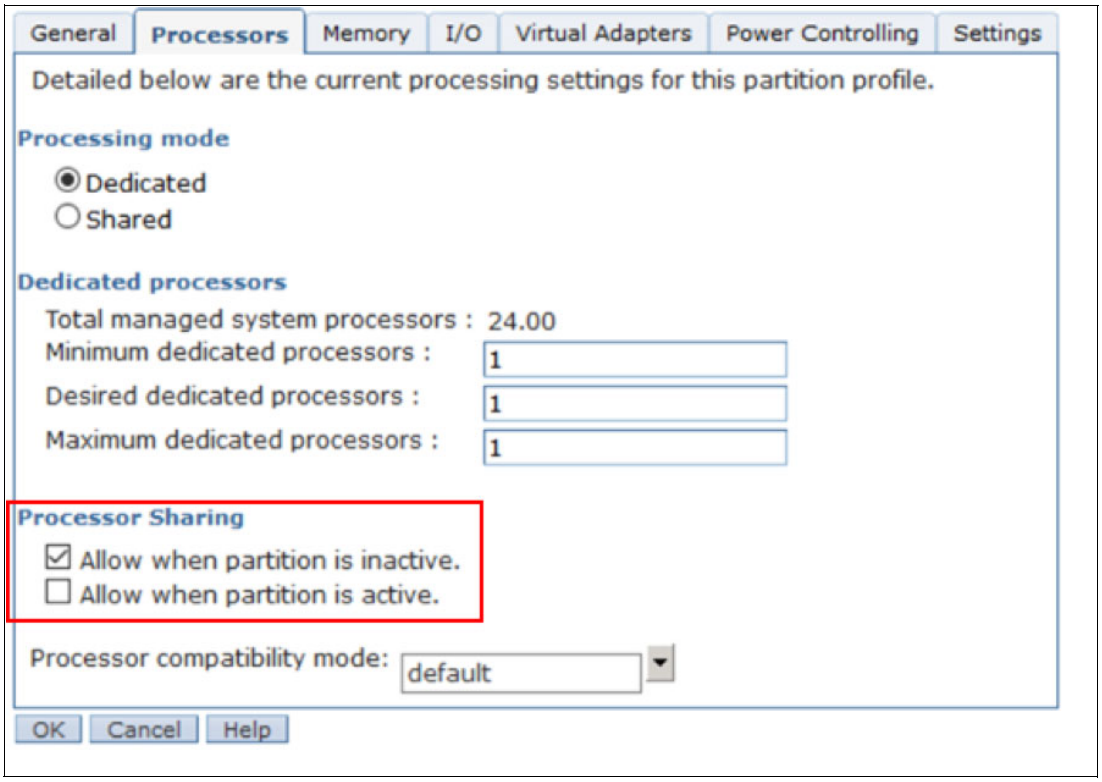


Figure 2-1 Partition profile window: Processing mode dedicated

You can specify to donate unused processor cycles only when the partition is not running (Figure 2-1) or when the partition is active.

Shared processor LPARs and Micro-Partitioning

SPLPARs are designed to maximize available physical processor resources. All processor resources that are not configured for dedicated, non-donating LPARs are available in the SPP. The processing capacity of an SPLPAR can be sized as small as 1/20th of a processor, with increments of 1/100th of a processor, allowing you to define more SPLPARs on a Power Systems server compared to dedicated LPARs. Besides the fine granularity in processor capacity sizing, SPLPARs do not exclusively own the processing capacity, which means that SPLPARs *always* donate unused capacity to the SPP. However, those LPARs can be configured as *uncapped* to use more processing capacity from the SPP when it is needed to fulfill peak load requirements, as shown in Figure 2-2.

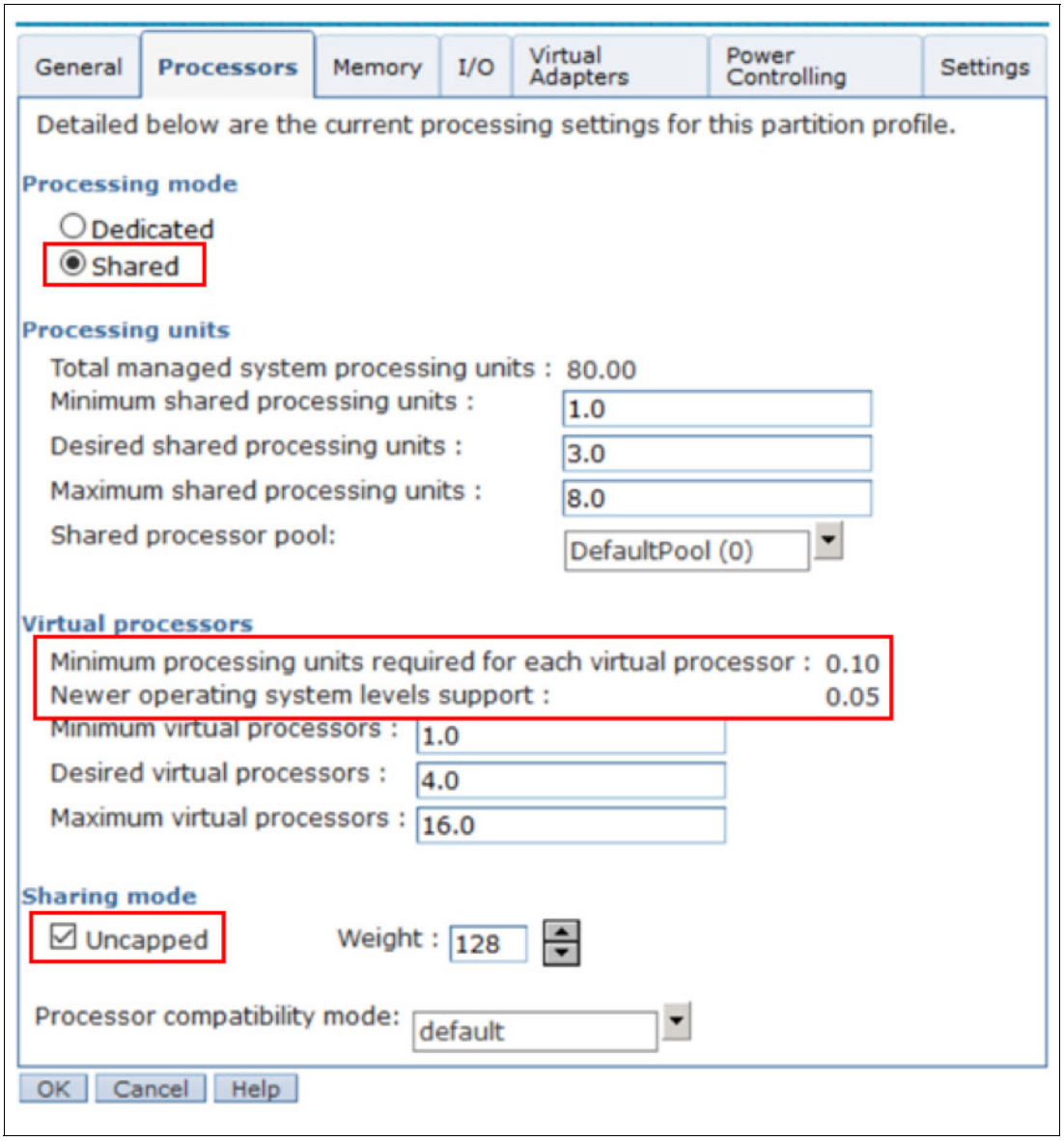


Figure 2-2 Partition profile window: Processing mode share

Shared processor pools

All processor resources that are not configured for dedicated LPARs are assigned to a global SPP. Initially, all shared processor resources belong to the DefaultPool, as shown in Figure 2-3, but can be configured in up to 64 SPPs. Those SPPs are subsets of and contained within the global SPP. Assigning SPLPARs to an SPP allows you to limit the processing capacity that is available to the SPLPAR, for example, to have control over license fees based on processor capacity.

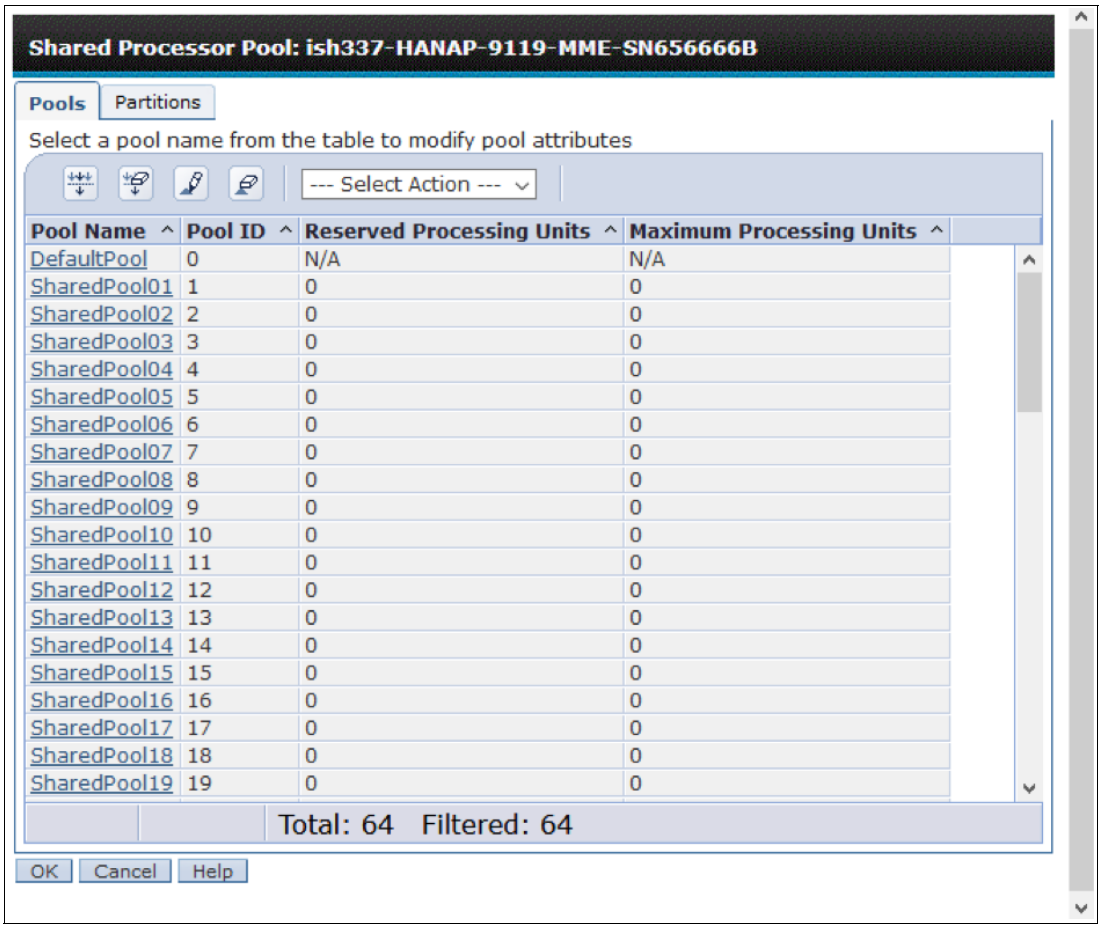


Figure 2-3 Shared Processor Pools window: Available pools

Simultaneous multithreading

Simultaneous multithreading (SMT) is a hardware feature of the POWER processor that provides the ability for a single physical processor to simultaneously dispatch instructions from more than one instruction thread context. POWER8 and POWER9 processor-based servers provide up to eight parallel threads (SMT8).

For these processors, when SMT mode is activated, a single physical processor appears to the OS as eight logical processors in the case of SMT8 or four logical processors in the case of SMT4, independent of the partition type. For example, in SMT4 mode, a partition with one dedicated physical processor operates with four logical processors. Similarly, in SMT8 mode, a shared processor partition with two virtual processors appear as a logical 16 processor partition. SMT is active by default.

SMT provides a capacity increase because it can interleave instructions from multiple instruction streams (run queues) on a single core. For example, process A stores a value, without SMT and process B's instruction must wait until the storage is complete. If the storage takes 5 cycles, 4 cycles go unused until process B can run. With SMT, process A's storage runs and is processed, and on the second processor cycle, process B's instruction runs. In essence, process B runs four cycles earlier than without SMT.

The performance improvement of SMT is application-dependent, but most commercial applications see a performance increase. For SAP environments, it is a best practice to enable SMT because the mixture of many parallel online users, RFC, and batch tasks benefits from this feature.

The PowerVM Management and Operations feature provides useful examples of HMC functions that simplify the operations to virtualize SAP landscapes on PowerVM.

Dynamic logical partition

DLPAR provides the capability to manually move resources (such as processors, memory, and I/O devices) to, from, and among running LPARs without shutting down or restarting the partitions. This feature was first introduced with POWER4 processor-based systems.

By allowing LPARs to redefine available system resources while online, DPLAR supports customers' needs to avoid potentially costly planned downtimes.

The following examples describe situations in which customers might consider employing DLPAR:

- Moving processors from a test partition to a production partition in periods of peak demand, then moving them back again as demand decreases.
- Moving memory to a partition that is doing excessive paging.
- Moving an infrequently used I/O device between partitions, such as a CD-ROM for installations or a tape drive for backups.
- Releasing a set of processor, memory, and I/O resources into the available pool so that a new partition can be created from those resources.
- Configuring a set of minimal LPARs to act as backup to primary LPARs and keeping some set of resources available.
- Assigning resources from a failed primary LPAR to a backup LPAR to continue workload processing.
- Temporarily assigning more capacity to an LPAR during an upgrade or migration to reduce SAP system downtime.

A DLPAR-safe program is one that does not fail as a result of DLPAR operations. Applications that are aware of the system topology can encounter problems if they do not expect the topology to be dynamic. For example, if the application is non-uniform memory access (NUMA)-aware but discovers the topology only at startup and NUMA nodes appear or disappear through DLPAR operations, unexpected results can occur. However, most programs are DLPAR-safe.

Note: For the support status of DLAR operations with SAP HANA, see SAP Note 2055470.

A program is DLPAR-aware if it anticipates that are system topology is dynamic and adjusts its use of resources. For example, to reduce NUMA effects, an application can create a pool of NUMA node-bound worker threads. DLPAR operations can increase or reduce the number of LPAR CPUs on NUMA nodes. So, the application can choose to adjust the number of worker threads in a NUMA node pool. A new NUMA node might also be introduced to the LPAR topology. In that case, the application might create a worker thread pool.

Dynamic Platform Optimizer

Introduced on IBM POWER7® processor-based servers, DPO is a PowerVM virtualization feature that aims to reduce NUMA effects by optimizing the placement of LPAR processor and memory resources on the server. By increasing processor-to-memory affinity and reducing the distances to memory, applications can experience performance improvements, depending on the workload.

LPAR topologies can be suboptimal (Figure 2-4) due to the following examples:

- Dynamic creation and deletion of partitions
- DLPAR operations that add and remove processors or memory
- Restarting with processor or memory configuration changes
- Hibernation (suspend or resume) operations
- LPM operations
- Central electronics complex (CEC) hot add, repair, and maintenance (CHARM)

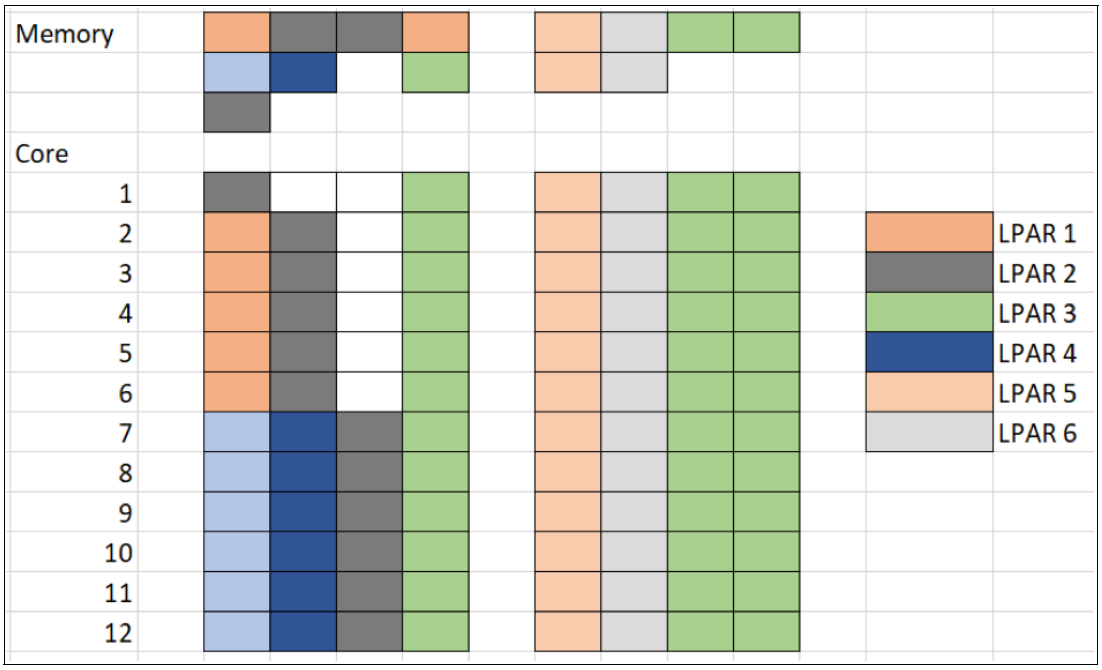


Figure 2-4 Suboptimal topology layout

Figure 2-4 shows some examples of suboptimal LPAR topologies that can be improved:

- LPAR 1 has processors and memory that is physically separated on different sockets.
- LPAR 2 has processors and memory that is allocated over multiple sockets.
- LPAR 3 has processors and memory that is allocated over multiple system nodes.

To assist with assessing the current topology, DPO can calculate an affinity score for each LPAR, which provides a quantified assessment of the LPAR layout ranging from 0 (poor) to 100 (best).

To improve the LPAR affinity, DPO determines an optimal resource placement strategy and runs a series of memory and processor relocations. For an estimate of improvement before running a DPO rebalance, a predicted affinity score can also be queried.

Figure 2-5 shows an improved topology after a DPO operation.

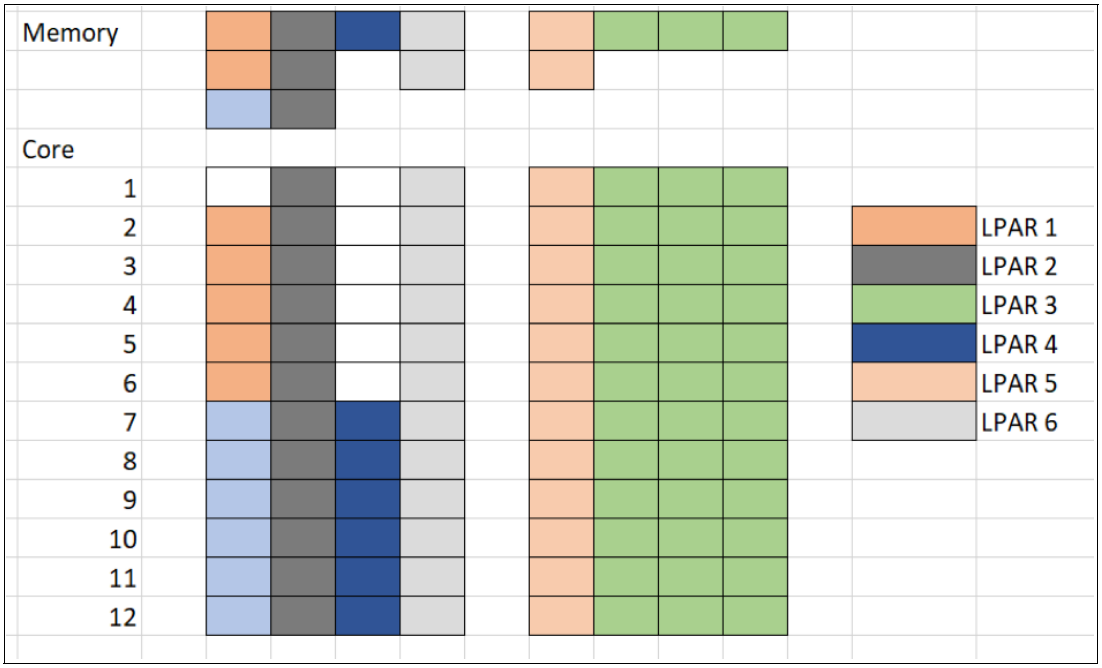


Figure 2-5 DPO optimized topology

DPO can be applied to individual LPARs or to the entire system. Specific LPARs can be listed and prioritized in a DPO operation. Likewise, LPARs can also be explicitly listed to be excluded from the operation. Those LPARs do not experience a topology change. LPARs that are not explicitly included or excluded are implicitly included but not prioritized for optimization, which means that those LPARs might be negatively impacted.

A DPO operation can be run on online and offline partitions. Online optimization allows for continued operation without planned downtimes. However, for applications that are NUMA-aware, both the OS and the application must react to a new LPAR topology to take advantage of the operation.

Note: For the support status of online DPO operations with SAP HANA, see SAP Note 2055470.

A DPO operation can take a significant amount of time, during which system performance is degraded and all LPARs can be impacted, including the ones that were excluded from operation.

2.3.2 Kernel-based virtual machine introduction

A KVM is a part of the open source virtualization infrastructure that turns the Linux kernel into an enterprise-class hypervisor. Quick Emulator (QEMU) is another part of this infrastructure. When the term KVM is used, it usually refers to the QEMU and KVM stack of software. To work efficiently, KVM requires hardware virtualization extensions. Hardware-assisted virtualization is a platform feature that enables the KVM hypervisor to take advantage of the underlying hardware when virtualizing guest OSs. Power Systems servers introduced virtualization assisted hardware with the POWER5 processor-based family of servers.

For more information, see *IBM PowerKVM: Configuration and Use*, SG24-8231.

KVM is composed of the following software stack:

- KVM performance
- QEMU
- IBM OpenPOWER Abstraction Layer

KVM performance

Because KVM is a thin layer over the firmware, it can deliver enterprise-grade performance to virtual machines (VMs) and consolidate a large amount of work on a single server. An important advantage of virtualization is the possibility of using resource overcommitment.

QEMU

QEMU is open source software that administers the VMs on a KVM hypervisor. It manages and monitors the VMs, and it performs the following basic operations:

- Create virtual image disks.
- Change the state of a VM:
  - Start a VM.
  - Stop a VM.
  - Suspend a VM.
  - Resume a VM.
  - Take and restore snapshots.
  - Delete a VM.
- Handle the I/O among guests and the hypervisor.
- Migrate VMs.

In a simplified view, the QEMU acts as the user space process for handling virtualization, and KVM acts as the kernel space module. QEMU can also work as an emulator, but this topic is beyond the scope of this publication.

IBM OpenPOWER Abstraction Layer

IBM OpenPOWER Abstraction Layer (OPAL) is a small layer of firmware that is available on POWER8 and POWER9 processor-based servers. It supports the KVM software stack and for partitions running without a hypervisor (bare metal). OPAL is part of the POWER firmware and provides an interface between the hardware and the KVM hypervisor. OPAL development is done by the public GitHub community. The code can be found at GitHub.

2.3.3 Resource overcommitment

Overcommitment is a mechanism to expose more CPU, I/O, and memory to the guest machine than exists on the real server, increasing server resource use and improving overall server utilization.

CPU virtualization

CPU virtualization is a technique that allows a virtual CPU to run on top of another CPU (virtual or physical). On Power Systems servers, the CPU virtualization of the guest instructions runs directly on the physical CPU, which reduces conversion overhead.

CPU overcommitment

CPU overcommitment allows an underutilized CPU to be shared with other VMs. The CPU overcommit is usually enabled when the VMs are not expected to use all of the CPU resources concurrently. Therefore, when one VM is not using its share of the CPU, another VM can use it. A CPU that is assigned to a VM is called a virtual central processing unit (vCPU). In an overcommitment scenario, the number of vCPUs is larger than the number of physical CPUs that is available.

Non-uniform memory access

NUMA describes an environment where memory access to different portions of memory can take significantly different amounts of time. In a NUMA environment, a processor usually has direct (local) access to memory that is connected to it. To access memory that is connected to a different but close processor, a longer data access path must be taken (remote memory). In the case of memory that is allocated from a processor in a different enclosure, an even longer path is taken (distant memory). The longer the path, the longer it takes to retrieve data from memory, so for best performance a KVM guest must use local memory. Within KVM, it is possible to define a NUMA environment for a guest. If a defined NUMA environment fits the physical architecture of the system, it results in increased performance. To link the vCPUs of a NUMA guest to the underlying physical CPU, use CPU pinning.

CPU pinning

CPU pinning allows the virtual CPUs of a guest VM to be *pinned* to a physical CPU or set of CPUs. The hypervisor schedules the work of a guest VM only on CPUs that the guest is pinned to (by default, the guest can be scheduled on any CPU). The advantage of pinning a VM is to reduce the time to access memory, and when the CPU's execution threads share data or instructions in the processors cache, the run time improves due to the reduced number of memory accesses.

CPU shares

In a KVM, the VMs run as processes on the host, which means that they are scheduled to run on host CPUs like any other process. The implication is that CPUs are shared by default. This CPU sharing allows CPU overcommitment, that is, creating more vCPUs than there are CPUs on the system. The Linux scheduler spreads the vCPUs among the CPU cores. However, when there is overcommitment, multiple vCPUs can share a CPU core. Shares are configured to balance the amount of time a VM is allotted compared to another VM.

Memory allocation

Guest memory is allocated by the host according to the guest configuration. It is possible to set a maximum amount of memory and a current amount. The guest has the maximum amount of memory that is available, but it can choose to use only the current amount and release the remaining amount to the host.

Memory ballooning

Memory ballooning is a technique that allows the guest memory to be increased or decreased cooperatively, depending on the amount of memory that is available on the guests and host. When memory ballooning is enabled on the guest, the hypervisor can remove and add memory to the guest dynamically.

This technique can be used if the memory is overcommitted, which means assigning the guests in sum more memory that the system provides. If a guest needs more memory and another guest needs less memory concurrently, the memory is used more efficiently. But, if all guests need their assigned overcommitted memory, this situation can degrade performance because the host starts swapping memory to disk.

2.3.4 Red Hat Virtualization

RHV is a virtualization management solution for virtualized servers and desktops. RHV enables central and effective management of the entire virtual environment, including virtual data centers; virtual clusters; hosts; guest servers and guest desktops; networking; and storage. The RHV hypervisor is based on KVM technology and supports VLANs, network bonding, and a wide range of network devices. RHV supports all storage systems that are certified on RHEL.

For more information, see Red Hat Virtualization.

RHV Manager introduction

RHV Manager (RHV-M) provides a centralized management system with a web-based GUI. The self-service user portal enables users to self-provision VMs, define templates, and administer their own environments.

Workload management

More memory and CPU resources can be added without disrupting applications. The built-in advanced service-level agreement (SLA) manager lets administrators define VM policies for CPU, memory, and network. These policies ensure quality of service (QoS). Administrators can also define per-user quotas for disk space, CPU usage, and memory.

Live migration

Guest live migration allows a seamless move of VMs from one host to another within an RHV cluster. Storage live migration allows for a single (or multiple concurrent) running VM disks to be moved within the storage infrastructure without interruption to users or the VM.

Resiliency, recovery, and maintenance

The high availability (HA) feature allows critical VMs to be restarted on another host in the event of hardware failure with three levels of priority. Snapshots allow cold or live snapshots to preserve a VM's current state. The maintenance mode allows one-click VM migration to put an RHV host into maintenance mode for upgrade or hardware updates. Affinity and anti-affinity workload grouping can be used to define workload affinity policies on how VMs run, either together on the same host or separately on different hosts.

3

IBM PowerVM management and operations

This chapter discusses PowerVM management and operation.

This chapter contains the following topics:

- Shared processor logical partitions
- Selecting and adjusting simultaneous multithreading
- Live Partition Mobility
- Dynamic Platform Optimizer and dynamic logical partition operations

3.1 Shared processor logical partitions

This section describes how to configure shared processor logical partitions (SPLPARs).

3.1.1 Configuring a shared processor LPAR

It is fairly simple to change a dedicated logical partition (LPAR) to an SPLPAR. All that you must do is make a small change in the existing profile followed by a restart of the LPAR. SPLPARs share their resources with other SPLPARs, but not all of these LPARs have the same importance to the business. If a *test* partition requests all shared resources, a production LPAR might suffer performance degradation and money is lost.

To protect the business, PowerVM supports the creation of a policy to protect running workloads. This policy determines the relative weights, guaranteed entitlements, and the number of virtual processors (virtual central processing units ((VCPUs)) for each LPAR.

The number of VCPUs is set so that it reflects the expected peak workload requirements. The entitlement (which is equal to the processing units) represents the minimum processing capacity that is allotted to an SPLPAR. The weight determines the relative priority among SPLPARS in the same pool. Examples are online versus batch, production versus non-production, and others.

Dynamic tools (dynamic logical partition (DLPAR) operations) can be used to fine-tune this policy during operation to address a load change. SAP monitoring reports and other monitoring tools can be used to determine the LPAR resource requirements for each SAP system.

A new profile or a change to an existing profile can be used to change an LPAR from dedicated to shared. The following steps show how to alter an existing profile, save it, and restart the partition:

1. Log in to the Hardware Management Console (HMC) and go to the LPAR. From the menu on the left, select **Partition Actions** → **Profiles** → **Manage Profiles** to access the Manage Profiles window.
2. Select the profile to be changed and select **Actions** → **Edit**. Alternatively, create a copy of the profile, and then change the new profile, as shown in Figure 3-1.

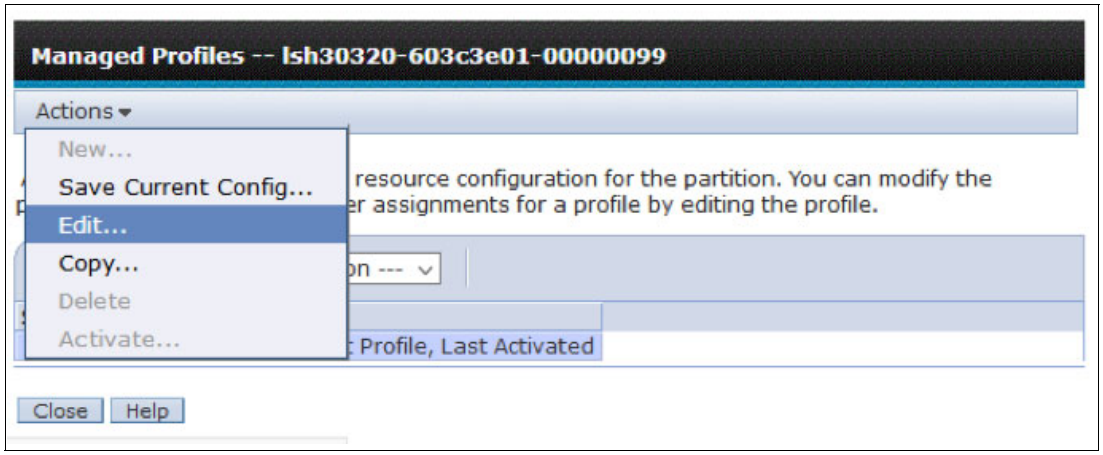


Figure 3-1 Managing Profiles window: Creating a copy of a profile

3. The profile opens, as shown in Figure 3-2. Select the **Processors** tab, select the **Shared** option, and change the resources according to your requirements.

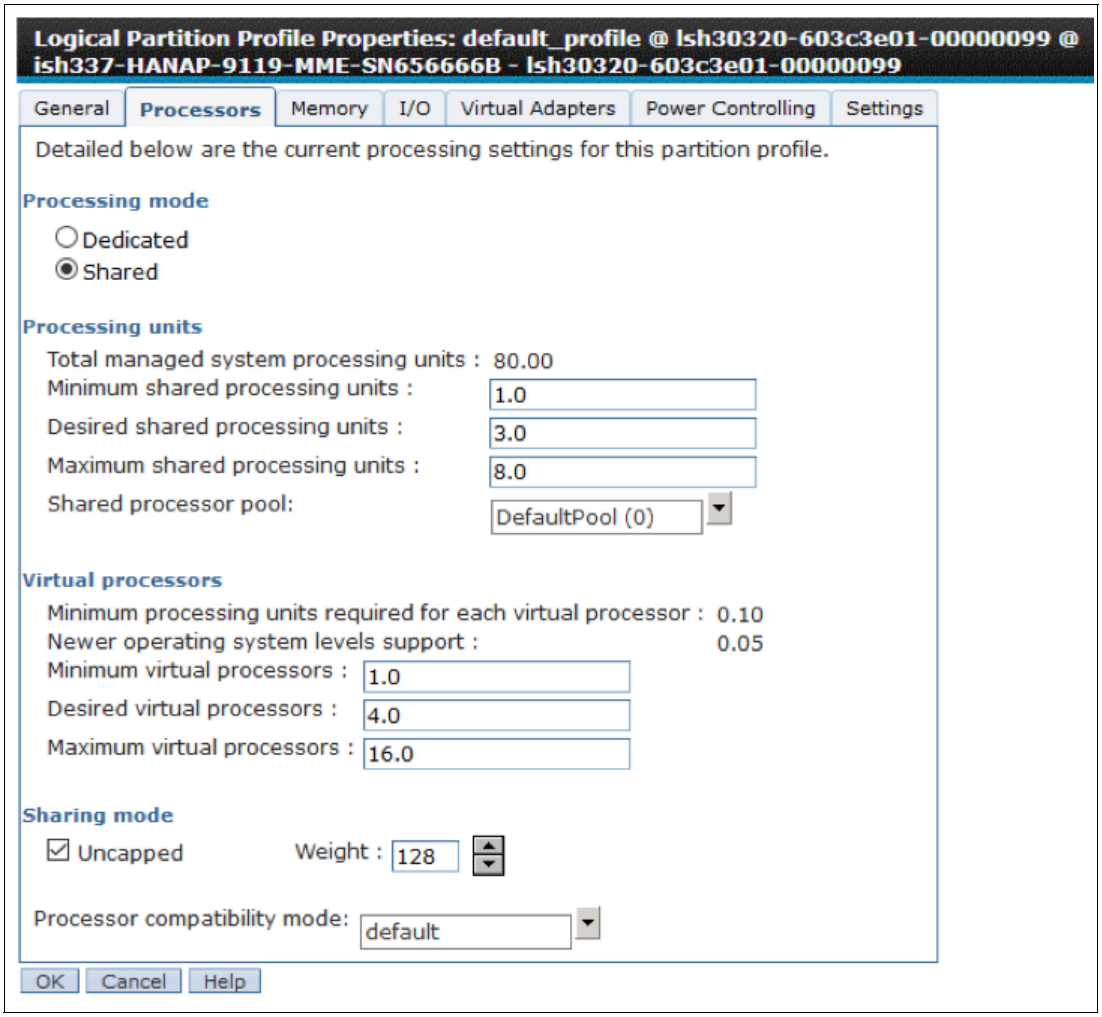


Figure 3-2 Processor window: Setting up the partition settings

Figure 3-2 shows the partition requesting three physical CPUs. The partition is guaranteed a minimum of one and a maximum of 8 physical CPUs. The operating system (OS) in the partition sees four CPUs, and DLPAR operations can reduce them to one or increase them to 16 CPUs. The LPAR is configured as Uncapped, which means that it can use more processing resources if it requires them and if free resources in the pool are available. Uncapped is the recommended sharing mode for SAP environments with PowerVM.

4. Shut down the LPAR and start it by activating the new or altered profile.

Micropartition design options

The SPLPARs come in two main types:

- Strict resource definitions
- Flexible limits

Strict resource definitions are achieved by defining the entitlement of the LPAR and capping the LPAR. The entitlement is the resource capacity that is guaranteed to the LPAR and is always available to it. Capping the LPAR restricts it from using any resources beyond this entitlement, so the chosen entitlement must be large enough to fulfill the processing requirements of the peak workload.

LPARs with strict limits are similar to dedicated LPARs with two exceptions:

- Other SPLPARs can use the resources when they are not used by this partition (the implementation is more flexible than dedicated donating partitions).
- Fractions of processors can be assigned, but dedicated LPARS always require whole processors.

LPAR entitlements act like reserved capacity: The sum of the entitlements of all the SPLPARs sharing the pool cannot exceed the number of physical processors in the pool so that the entitlements can be guaranteed. Therefore, large SPLPAR entitlements can limit the number of SPLPARs that are active in the shared processor pool (SPP).

To achieve the best processor sharing, keep the entitlements small and the partitions uncapped. Uncapped partitions can use resources in the shared pool beyond their own guaranteed entitlement. This is the most common approach for combined SAP landscapes because the SPLPARs can adapt to the workload requirements and the resource distribution can adapt to the changing workload patterns in multiple SPLPARs. The SPP can experience contention if workloads are heavily competing for resources.

There are several options to enforce a priority policy and relative importance of the SPLPARs, which are referred to as *implicit capping*. They provide the means to control the resource allocation, although still providing the flexibility and efficiency of resource sharing.

Implicit capping of SPLPARs

Most LPAR settings can be dynamically altered during run time by DLPAR operations that use the HMC. Using this function, the behavior of the LPARs can be changed.

The number of VCPUs that are available to a partition limit the number of whole physical processors that this LPAR can use because one vCPU represents at maximum one physical processor. The LPAR is effectively *capped* at a maximum physical processor utilization that is equal to its number of configured VCPUs.

By reducing the SPLPAR weight, the LPAR is not capped; only its ability to compete for more resources over its actual entitlement is reduced. If there is no competition for these resources, the SPLPAR is restricted only by its number of configured VCPUs, which allows it to be both flexible and not harmful to other workloads with higher priority (greater weights).

The advantages of implicit capping versus explicit capping are:

- The flexibility that is gained by low reserved entitlement, which allows more SPLPARs to be defined on the same number of physical processors versus a second *implicit entitlement* that can be guaranteed.
- SPLPARs with low weights can be protected from uncapped SPLPAR with higher weights using all the resources.

- You can limit the maximum processors in hosting environments.
- Implicit capping can be adjusted by using dynamic LPAR operations.

Note: A partition that is capped and a partition that is uncapped with a weight of 0 are functionally identical. In terms of performance, the results are the same because the processor utilization can go up to only the partition entitled capacity. The monitoring of an SAP system can be impacted because a partition shows up as uncapped but behaves like a capped partition. When monitoring an uncapped SPLPAR, always check the weight value.

Managing multiple SPPs

Multiple SPPs are useful in SAP landscapes in several ways, including the ability to separate workloads and to create subpools, which can be used to limit software license costs when they are based on the number of available processors.

Complete the following steps:

1. To configure an SPP, click **Shared Processor Pool** from the HMC task list on the left, as shown in Figure 3-3. A new window opens and shows the configuration of the available SPPs.

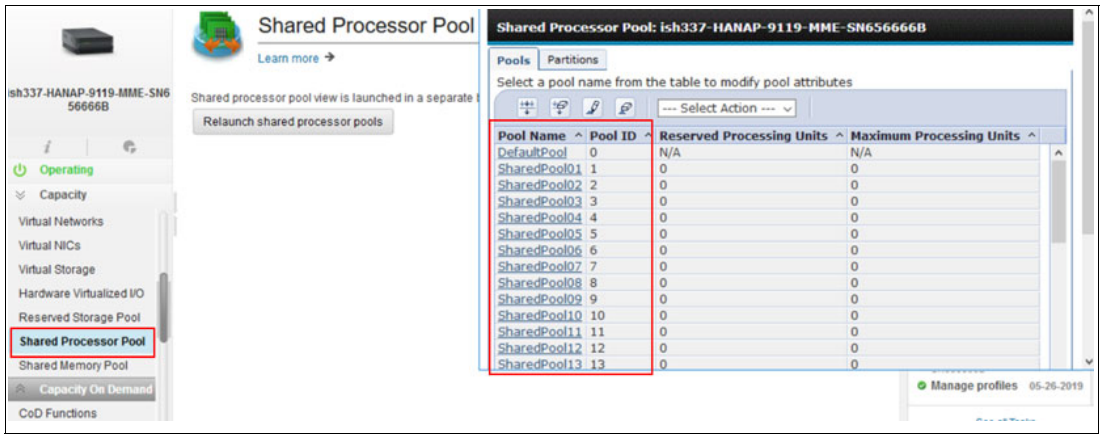


Figure 3-3 Configuring the shared processor pool

2. As shown in Figure 3-3 on page 23, initially all processors (processing units) are available in the DefaultPool. To create another SPP, for example, SharedPool01, processing units must be assigned, as shown in Figure 3-4.

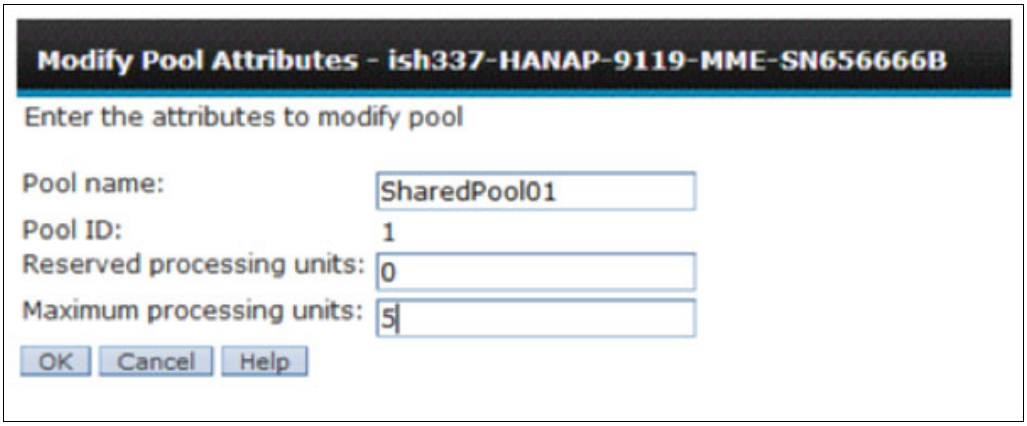
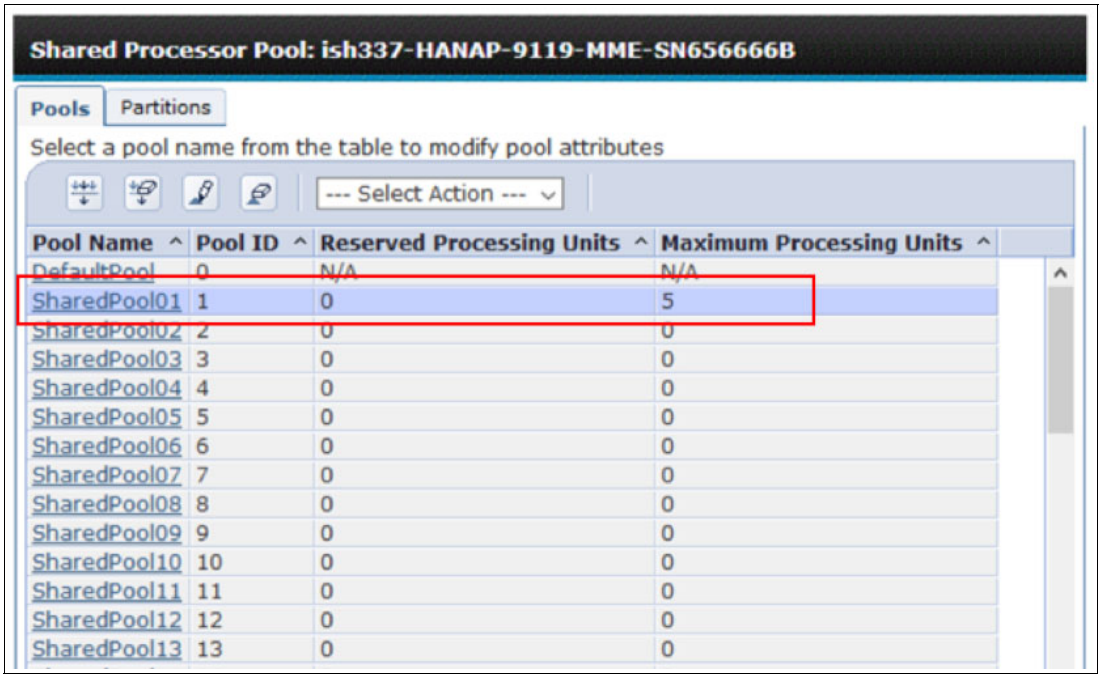


Figure 3-4 Creating a shared processor pool

3. Click **OK**. Figure 3-5 shows the new SPP that is named SharePool01.



Pool Name	Pool ID	Reserved Processing Units	Maximum Processing Units
DefaultPool	0	N/A	N/A
SharedPool01	1	0	5
SharedPool02	2	0	0
SharedPool03	3	0	0
SharedPool04	4	0	0
SharedPool05	5	0	0
SharedPool06	6	0	0
SharedPool07	7	0	0
SharedPool08	8	0	0
SharedPool09	9	0	0
SharedPool10	10	0	0
SharedPool11	11	0	0
SharedPool12	12	0	0
SharedPool13	13	0	0

Figure 3-5 Showing the new shared processor pool

4. After the SPP is created, a partition can be assigned to the pool by using the same interface. Click the **Partitions** tab and select the partition name, as shown in Figure 3-6 on page 25.

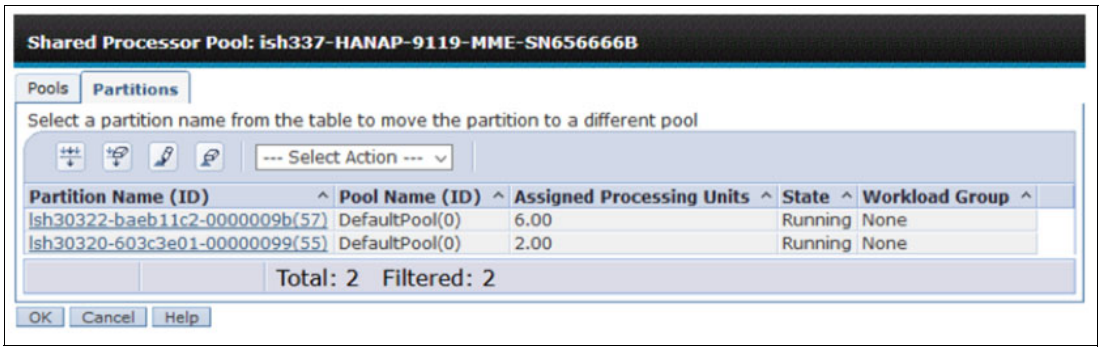


Figure 3-6 Assigning a shared processor pool to a partition

5. Select the SPP to be assigned to this partition, as shown in Figure 3-7.

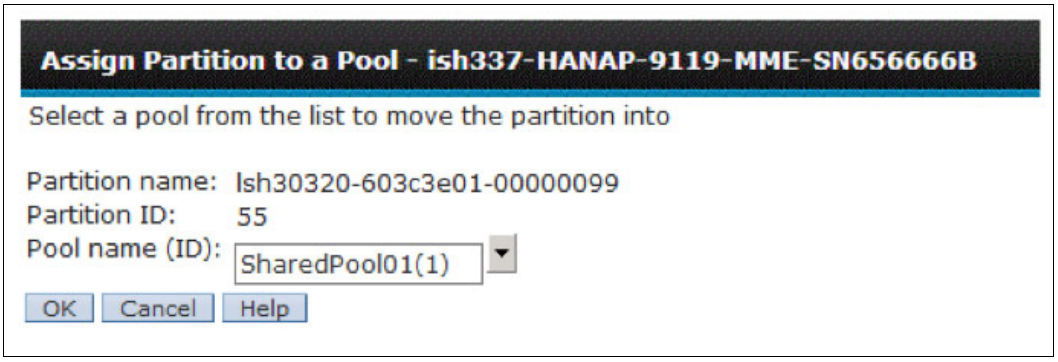


Figure 3-7 Selecting the shared processor pool to assign to the partition

The partition can now use only the processor capacity that is available in SharedPool01, as shown in Figure 3-8.

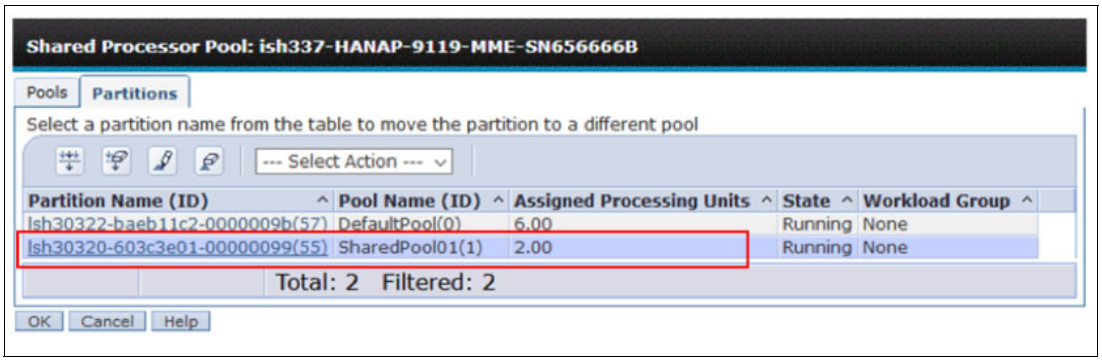


Figure 3-8 Shared processor pool resources for the partition

Enabling the pool utilization authority

In an SPP environment, the pool utilization authority (PUA) can be used to restrict the visibility of the monitoring data of an LPAR, which can be useful, for example, in hosted environments where customers cannot collect information about the hosting system. However, monitoring processor usage in SAP systems likely requires more metrics from the SPP.

The PUA is configured separately for each partition. To explicitly grant the PUA to an SPLPAR, click **Enable Performance Information Collection** on the HMC.

Select the SPLPAR in the HMC and open the **General Properties** page. Click **Advanced** to show the advanced properties of the partition and scroll down. Click **Enable Performance Information Collection**, and click **Save** to immediately activate the change, as shown in Figure 3-9.

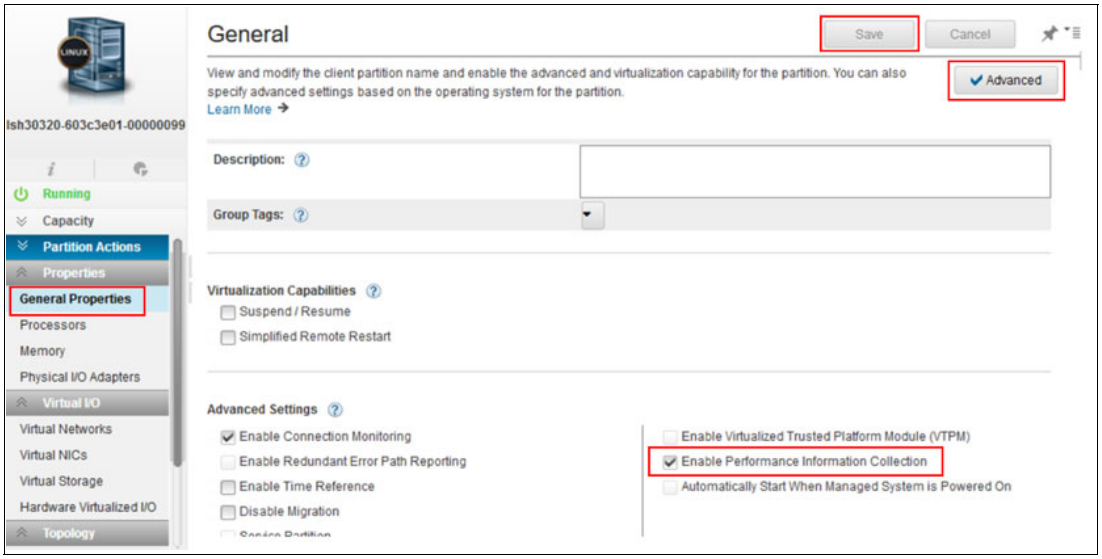


Figure 3-9 General window: General Properties section

3.2 Selecting and adjusting simultaneous multithreading

Simultaneous multithreading (SMT) was introduced as two-way multithreading (SMT-2) with POWER5 processors, followed by SMT4 in POWER7 processors and SMT8 in POWER8 processors.

With POWER9 processors, most workloads perform best when using SMT-8, which is also the default when a Linux LPAR starts. There are a few single-threaded applications that benefit from lower SMT levels.

The SMT level is set with OS commands. It can be turned on or off, and the number of hardware threads can be set explicitly. Changes to the SMT mode affect the whole partition and take effect immediately.

After a restart, a Linux LPAR always defaults to the maximum supported SMT level regardless of the SMT level that was active before the LPAR shut down. If the LPAR must run with a lower SMT level, then it must be set manually upon each restart, for example, within a **`boot.local`** startup script.

The SMT mode of a Linux LPAR is managed by the **`ppc64_cpu`** command that uses the parameter **`--smt`**. The SMT mode can be set to either on or off, and it can be set directly.

To turn on SMT, run one of the following commands:

- **`# ppc64_cpu --smt=on`**
- **`# ppc64_cpu --smt=8`**

To turn off SMT, run one of the following commands:

- **`# ppc64_cpu --smt=off`**
- **`# ppc64_cpu --smt=1`**

To select SMT-4 explicitly, run the following command:

```
# ppc64_cpu --smt=4
```

The current SMT setting is displayed by running the following command:

```
# ppc64_cpu --smt
SMT=4
```

3.3 Live Partition Mobility

Live Partition Mobility (LPM) is a component of PowerVM Enterprise Edition that can move AIX, IBM i, and Linux LPARs from one physical system to another one.

LPM is used in SAP Landscapes to reduce planned downtime. SAP includes LPM operations into their products, such as SAP Landscape Management (SAP LaMa) to take advantage of the mobility capability.

The mobility process transfers the system environment, including the processor state, memory, attached virtual devices, and connected users. When the mobility process is ongoing, users of an LPAR are not affected and are not exposed to what is happening in the background. Often, LPM is considered when the administrator of the Power Systems servers does maintenance of the system, performs a planned outage, or sets resources.

When performing LPM, you must have another Power Systems server that is configured almost the same as the initial Power Systems server in terms of network and storage. Errors in setup or connectivity can cause the LPM to fail.

Here are a few prerequisites that must be considered when performing an LPM.

Note: The destination firmware level must be the same as the source firmware level, or later if compatible. To check for compatibility, see IBM Knowledge Center.

- Check the logical memory block (LMB) matches on both systems. Log in to the HMC and use the following command:

  ```
  lshwres -r mem -m <machine name> --level sys -F mem_region_size
  ```

- Check that the destination system has enough memory to support the mobile partition.
- Check that the destination system has enough processors to support the mobile partition.
- Check the source and destination Virtual I/O Server (VIOS) levels.
- Check that the mover service partition (MSP) attribute is selected on at least one VIOS in both systems.
- Check that Shared Ethernet Adapter (SEA) VLAN is used for the LPAR to be moved and present on the destination system.
- Check that all disks:
  - If mapped by using VSCSI, are shared with one of the VIOSs at the destination.
  - If using N-Port ID Virtualization (NPIV), the WWPN for virtual client adapter is both masked and zoned properly on the switch and the storage.
- LPARs must not have any physical hardware that is attached to them. All the storage and network connections to the LPAR must be virtual.

For more information about the checklist, see the following resources:

- Live Partition Mobility Setup Checklist
- *IBM Power Systems: Live Partition Mobility*

3.3.1 Types of Live Partition Mobility

There are two types of LPM that can be performed on an LPAR:

- Active partition mobility
- Inactive partition mobility

Active partition mobility is performed on a running LPAR. Running applications on the LPAR are not interrupted, but all the resources of the LPAR are moved to a different server.

Inactive partition mobility moves an LPAR that is in *not* in an activated state. It is often used to preserve LPAR features like storage, OS images, data disks, and the partition profile, which can be used to activate the LPAR again. This option is used to decongest a Power Systems server.

When LPM is performed, the minimum requirement of migrating the LPAR must be checked. Before LPM is performed, a validation check is run, which checks for storage, network, OS level, and other aspects that can result in LPM failure. Before migrating an LPAR, run a validation, and if errors occur they must be corrected. Otherwise, the migration might fail and in extreme cases LPAR data might be corrupted.

Installations of SAP HANA can use LPM if the target infrastructure is supported and all prerequisites from SAP Note 2055470 are fulfilled. Before the LPM process is run, SAP HANA must be stopped. Perform a valid full backup before moving LPARs among host systems and follow all recommendations of the solution vendor.

SAP HANA must be stopped because the LPAR placement can *change* after an LPM operation. For more information about LPAR placement, see “Non-uniform memory access” on page 16.

Concerns about functional issues can arise after LPM is performed on an active SAP HANA database. The non-uniform memory access (NUMA) placement of an LPAR is likely to change after migration. The SAP HANA database is a NUMA-aware application, so SAP requires a restart of SAP HANA in production systems after LPM completes to adapt to the new topology.

Before attempting LPM on a Linux LPAR, ensure that the Linux version is at least the minimum supported level for LPM on Power Systems:

- SUSE Linux V10 SP1
- Red Hat Linux Version 5.1

The following sections demonstrate LPM for an SAP HANA LPAR. Screen captures from the HMC console are used to visualize the process.

Throughout this demonstration, the name of the LPAR to be migrated by using LPM is lsh30021-3ac0eacd-000000f6. This is the name of the managed system on the HMC. Its host name on OS level is lsh30021, which is used as a short name going forward. As shown in Figure 3-10 on page 29, the LPAR lsh30021 allocated 12 CPUs and 1000 GB memory and is in a running state. An active LPM is performed, and to run an inactive LPM, the LPAR must be shut down first.

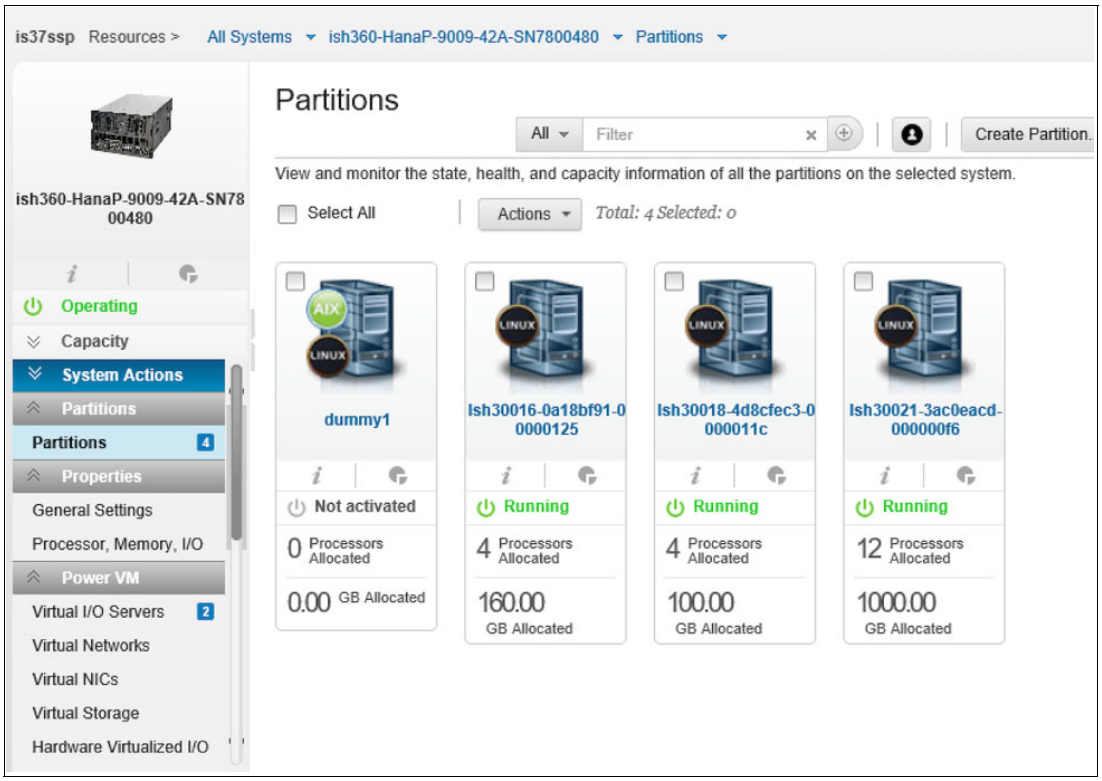


Figure 3-10 Partitions window

Figure 3-11 shows a snapshot of the properties of lsh30021. It is connected to the network by using a SEA, and all attached disks use NPIV. The OS is SUSE Linux SLES 12 SP3.

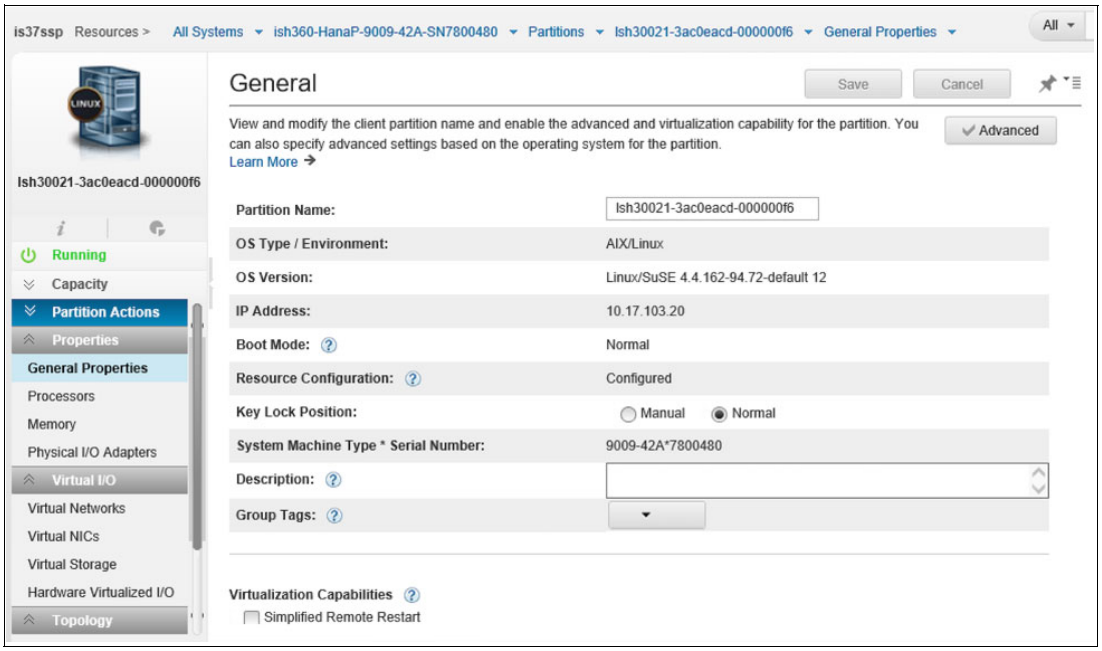


Figure 3-11 Snapshot of partition lsh30021

Figure 3-12 shows all virtual networks that are associated with the selected LPAR.

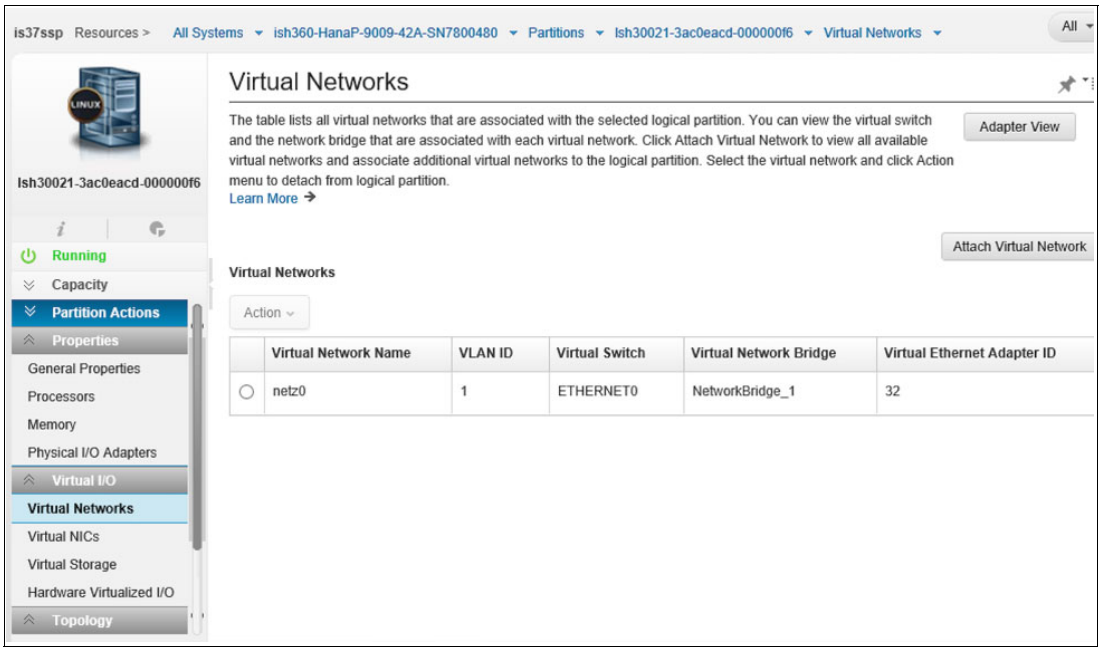


Figure 3-12 Virtual Networks table

Figure 3-13 shows the virtual storage channels that are assigned to the PowerVM configuration.

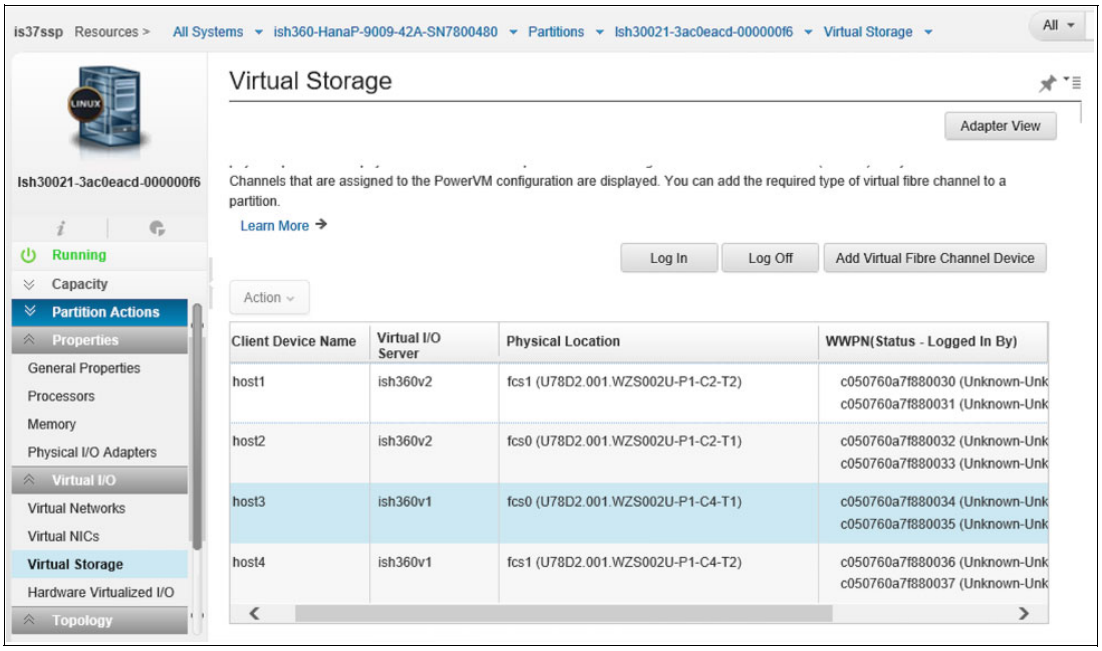


Figure 3-13 Virtual storage that is assigned to the PowerVM configuration

Non-uniform memory access and SAP HANA status

Example 3-1 on page 31 shows the output of **`numactl --hardware`**, which documents the NUMA topology. The LPAR has 96 logical CPUs (SMT threads) and 1000 GB of memory, all on node 1. The same command is used to check the NUMA topology after the LPM is complete.

Example 3-1 Output of the current NUMA topology

```
[root@lsh30021 ~]# numactl --hardware
available: 2 nodes (0-1)
node 0 cpus:
node 0 size: 0 MB
node 0 free: 0 MB
node 1 cpus: 0 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28
29 30 31 32 33 34 35 36 37 38 39 40 41 42 43 44 45 46 47 48 49 50 51 52 53 54 55 56 57 58
59 60 61 62 63 64 65 66 67 68 69 70 71 72 73 74 75 76 77 78 79 80 81 82 83 84 85 86 87 88
89 90 91 92 93 94 95
node 1 size: 1022928 MB
node 1 free: 774444 MB
node distances:
node   0   1
  0:  10  40
  1:  40  10
[root@lsh30021 ~]# su - db1adm

db1adm@lsh30021:/usr/sap/DB1/HDB00> HDB info
USER          PID     PPID  %CPU        VSZ        RSS COMMAND
db1adm      21902    21901   0.2     112960       5312 -sh
db1adm      21982    21902   0.0     112064       4096  \_ /bin/sh /usr/sap/DB1/HDB00/HDB
info
db1adm      22013    21982   0.0     118784       8576      \_ ps fx -U db1adm -o
user:8,pid:8,ppid:8,pcpu:5,vsz:10,rss:10,args
db1adm      23984        1   0.0      10048       4928 sapstart
pf=/hana/shared/DB1/profile/DB1_HDB00_lsh30021.sap.corp
db1adm      23993    23984   0.0     411456      77824  \_
/usr/sap/DB1/HDB00/lsh30021.sap.corp/trace/hdb.sapDB1_HDB00 -d -nw -f
/usr/sap/DB1/HDB00/lsh30021.sap.corp/daemon.ini pf=/usr/sap/DB1/SY
db1adm      24011    23993   1.3   12814912    8407040      \_ hdbnameserver
db1adm      24296    23993   0.3    4630464     450432      \_ hdbcompileserver
db1adm      24299    23993   0.4    8646528    6714688      \_ hdbpreprocessor
db1adm      24346    23993  11.2  163612352  160016384      \_ hdbindexserver -port 30003
db1adm      24349    23993   0.9    7534400    1728448      \_ hdbxsengine -port 30007
db1adm      25092    23993   0.3    6098112     689088      \_ hdbwebdispatcher
db1adm      18094        1   0.0    6894656      46080 hdbrsutil  --start --port 30003
--volume 3 --volumesuffix mnt00001/hdb00003.00003 --identifier 1573188216
db1adm      17518        1   0.0     668672      40448 hdbrsutil  --start --port 30001
--volume 1 --volumesuffix mnt00001/hdb00001 --identifier 1573188180
db1adm      17139        1   0.0     545024      39744 /usr/sap/DB1/HDB00/exe/sapstartsrv
pf=/hana/shared/DB1/profile/DB1_HDB00_lsh30021.sap.corp -D -u db1adm
db1adm@lsh30021:/usr/sap/DB1/HDB00>
```

The commands that are illustrated in Example 3-2 indicate that the SAP HANA database is running.

Example 3-2 Checking the status of the SAP HANA database

```
db1adm@lsh30021:/usr/sap/DB1/HDB00> sapcontrol -nr 00 -function GetProcessList

24.05.2019 08:25:09
GetProcessList
OK
name, description, dispstatus, textstatus, starttime, elapsedtime, pid
hdbdaemon, HDB Daemon, GREEN, Running, 2019 04 24 05:54:07, 1010:31:02, 23993
hdbcompileserver, HDB Compileserver, GREEN, Running, 2019 04 24 05:54:16, 1010:30:53, 24296
hdbindexserver, HDB Indexserver-DB1, GREEN, Running, 2019 04 24 05:54:16, 1010:30:53, 24346
hdbnameserver, HDB Nameserver, GREEN, Running, 2019 04 24 05:54:07, 1010:31:02, 24011
hdbpreprocessor, HDB Preprocessor, GREEN, Running, 2019 04 24 05:54:16, 1010:30:53, 24299
```

```
hdbwebdispatcher, HDB Web Dispatcher, GREEN, Running, 2019 04 24 05:55:11, 1010:29:58,
25092
hdbxsengine, HDB XSEngine-DB1, GREEN, Running, 2019 04 24 05:54:16, 1010:30:53, 24349
db1adm@lsh30021:/usr/sap/DB1/HDB00>
```

SAP HANA is now stopped in preparation for LPM, as shown in Example 3-3.

Example 3-3 Stopping SAP HANA

```
db1adm@lsh30021:/usr/sap/DB1/HDB00> HDB stop
hdbdaemon will wait maximal 300 seconds for NewDB services finishing.
Stopping instance using: /usr/sap/DB1/SYS/exe/hdb/sapcontrol -prot NI_HTTP -nr 00 -function Stop 400

24.05.2019 08:40:43
Stop
OK
Waiting for stopped instance using: /usr/sap/DB1/SYS/exe/hdb/sapcontrol -prot NI_HTTP -nr 00 -function
WaitforStopped 600 2


24.05.2019 08:41:51
WaitforStopped
OK
hdbdaemon is stopped.
```

Example 3-4 and Example 3-5 show performing the checks to verify that SAP HANA stopped.

Example 3-4 Confirming that SAP HANA stopped

```
db1adm@lsh30021:/usr/sap/DB1/HDB00> HDB info
USER          PID     PPID  %CPU        VSZ        RSS COMMAND
db1adm      21902    21901   0.0     112960       5440 -sh
db1adm      30903    21902   0.0     112064       4096  \_ /bin/sh /usr/sap/DB1/HDB00/HDB info
db1adm      30934    30903   0.0     118784       8512      \_ ps fx -U db1adm -o
user:8,pid:8,ppid:8,pcpu:5,vsz:10,rss:10,args
db1adm      18094        1   0.0    6894656      46080 hdbrsutil  --start --port 30003 --volume 3
--volumesuffix mnt00001/hdb00003.00003 --identifier 1573188216
db1adm      17518        1   0.0     668672      40448 hdbrsutil  --start --port 30001 --volume 1
--volumesuffix mnt00001/hdb00001 --identifier 1573188180
db1adm      17139        1   0.0     610560      39872 /usr/sap/DB1/HDB00/exe/sapstartsrv
pf=/hana/shared/DB1/profile/DB1_HDB00_lsh30021.sap.corp -D -u db1adm
db1adm@lsh30021:/usr/sap/DB1/HDB00> HDB proc
USER       PID  PPID %CPU    VSZ   RSS COMMAND
db1adm   30949 21902  0.0 112064  4160  \_ /bin/sh /usr/sap/DB1/HDB00/HDB proc
db1adm   30982 30949  0.0 112064  1024      \_ /bin/sh /usr/sap/DB1/HDB00/HDB proc
db1adm   18094     1  0.0 6894656 46080 hdbrsutil  --start --port 30003 --volume 3 --volumesuffix
mnt00001/hdb00003.00003 --identifier 1573188216
db1adm   17518     1  0.0 668672 40448 hdbrsutil  --start --port 30001 --volume 1 --volumesuffix
mnt00001/hdb00001 --identifier 1573188180
db1adm   17139     1  0.0 610560 39872 /usr/sap/DB1/HDB00/exe/sapstartsrv
pf=/hana/shared/DB1/profile/DB1_HDB00_lsh30021.sap.corp -D -u db1adm
db1adm@lsh30021:/usr/sap/DB1/HDB00>
```

Example 3-5 More checks to confirm that SAP HANA stopped

```
db1adm@lsh30021:/usr/sap/DB1/HDB00> sapcontrol -nr 00 -function GetProcessList

24.05.2019 08:42:57
GetProcessList
OK
name, description, dispstatus, textstatus, starttime, elapsedtime, pid
```

```
hdbdaemon, HDB Daemon, GRAY, Stopped, , , 23993
db1adm@lsh30021:/usr/sap/DB1/HDB00>
```

After confirming SAP HANA stopped, an LPM migration is performed by using the HMC GUI:

1. Select the partition to be migrated, and then select the **Migrate** option in the Partition Actions menu.

   The Partition Migration dialog opens at the Migration Information step, as shown in Figure 3-14. Each step asks for migration-relevant information. The default is often sufficient, but you can always change it.

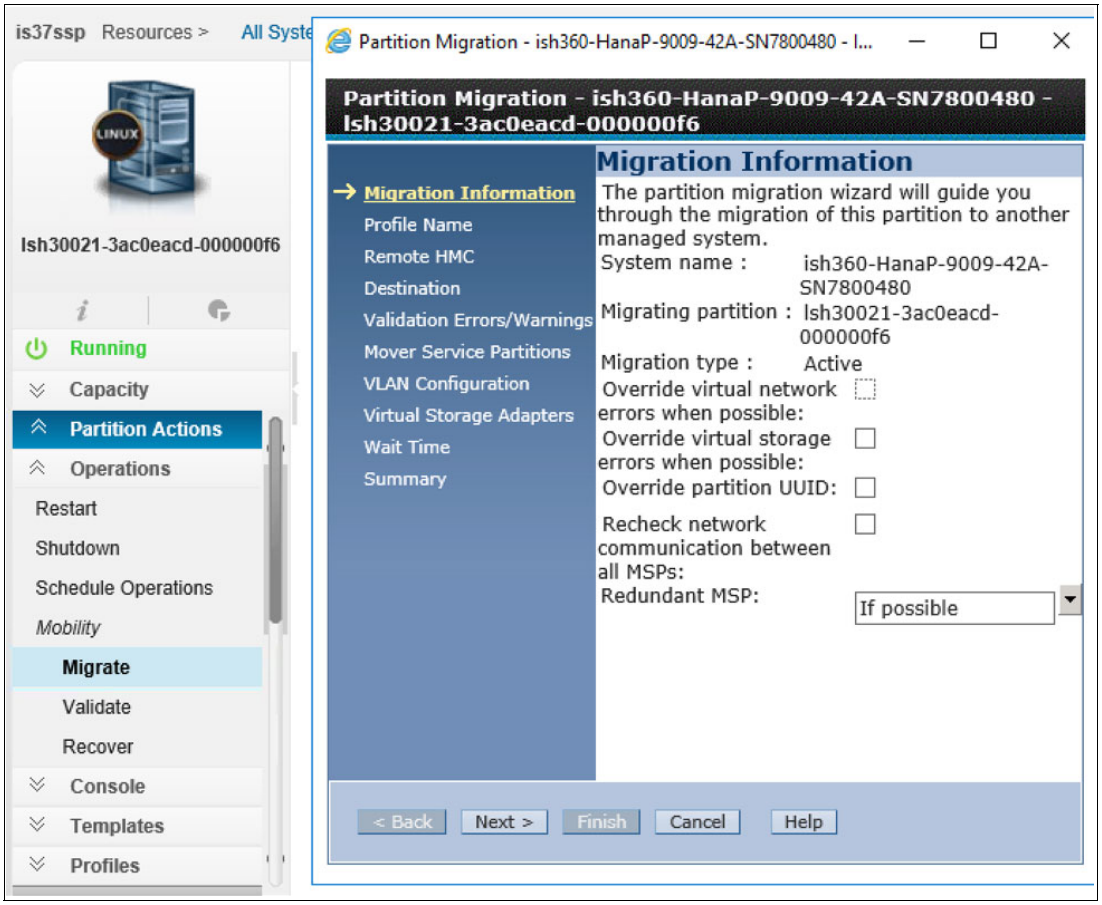


Figure 3-14 Partition Actions: Migration Information

Figure 3-15 shows the Profile Name window to create a migration profile with the current state of the partition.

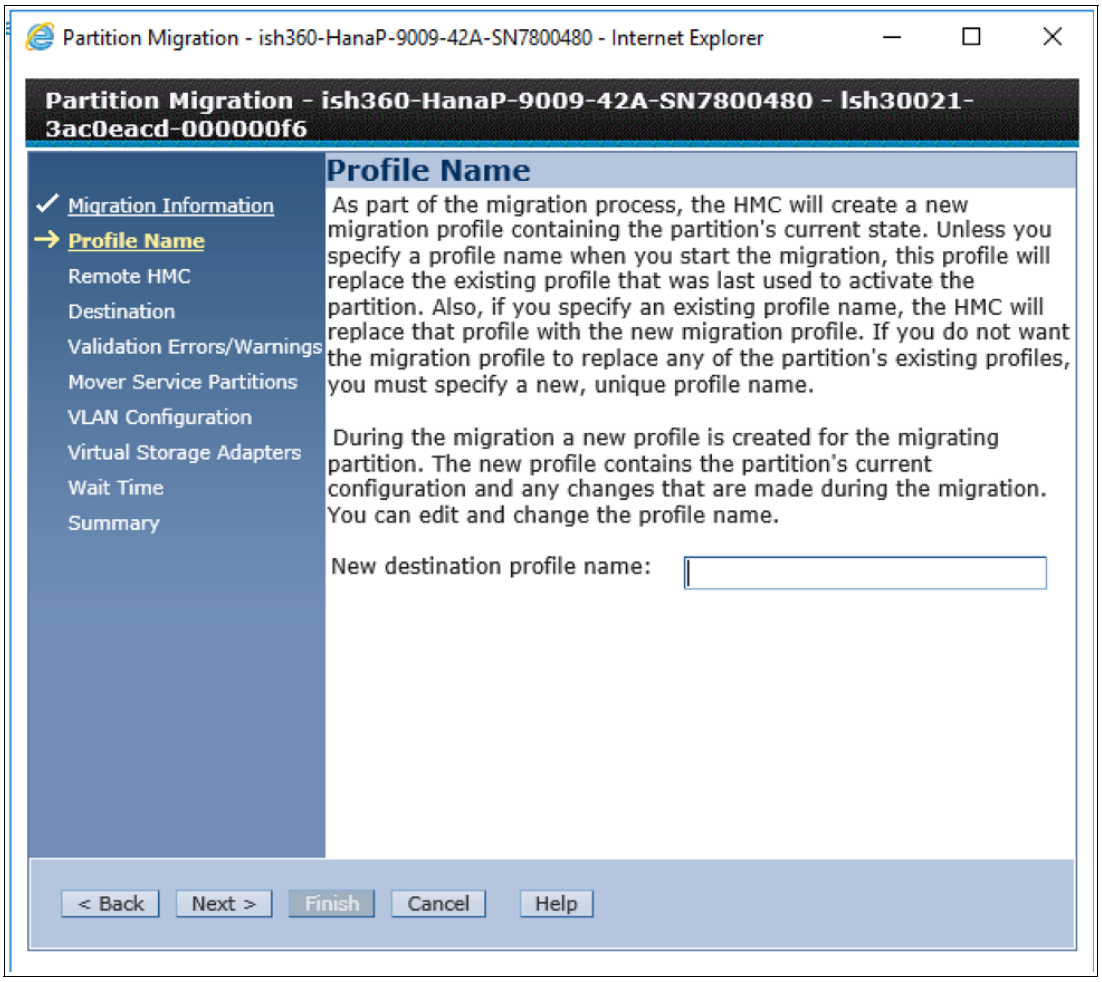


Figure 3-15 Profile Name window: Migration profile

2. Figure 3-16 shows that one HMC is managing the source and the destination system, so the **Remote Migration** check box in the Remote HMC step stays clear. If the destination system is managed by a different HMC, the destination HMC information must be entered in this step.

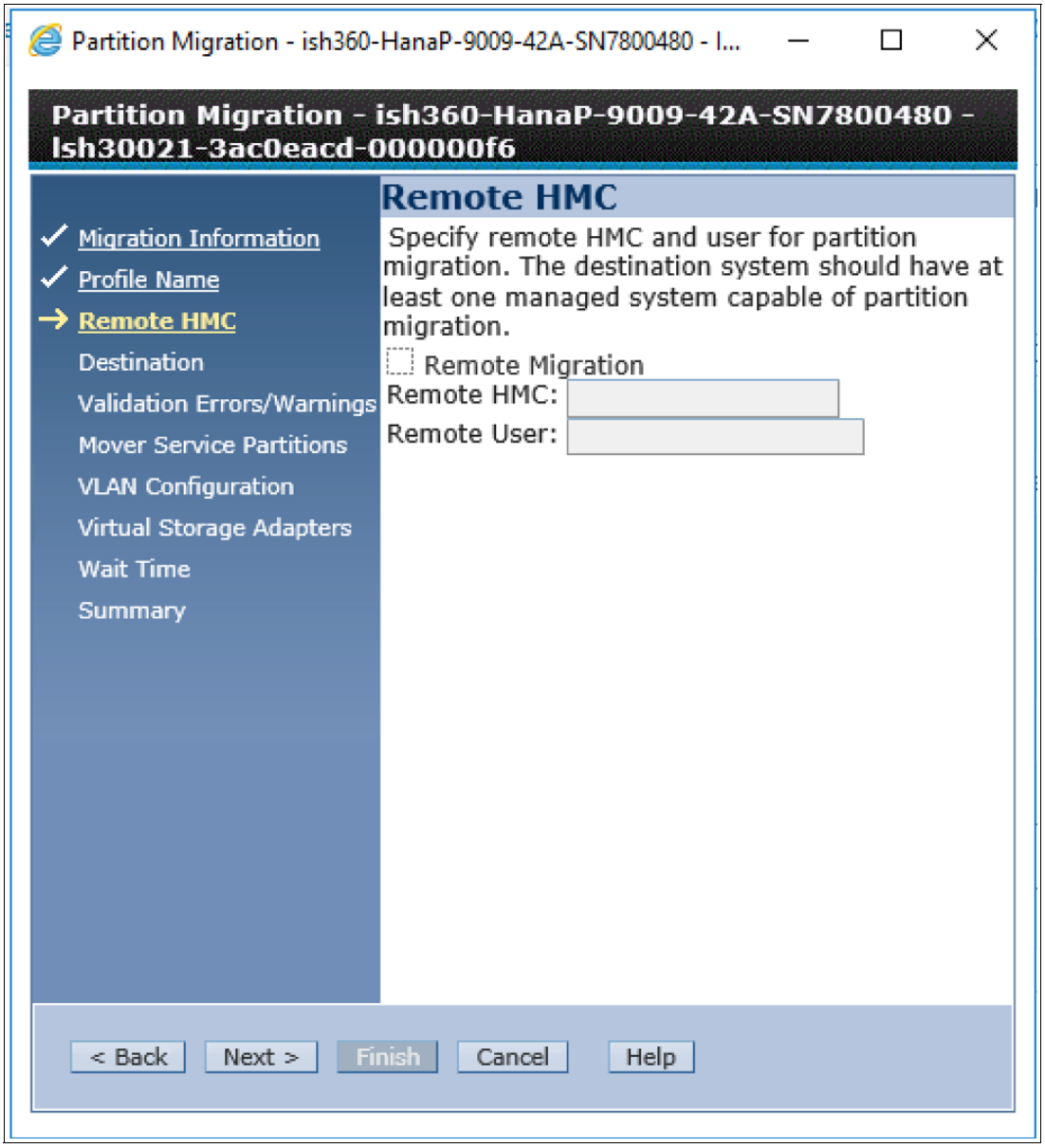


Figure 3-16 Remote HMC window: Specifying the remote HMC and user for partition migration

3. Select the destination system in the Destination step, as shown in Figure 3-17. It can be selected from a drop-down list, which is supplied by the destination HMC.

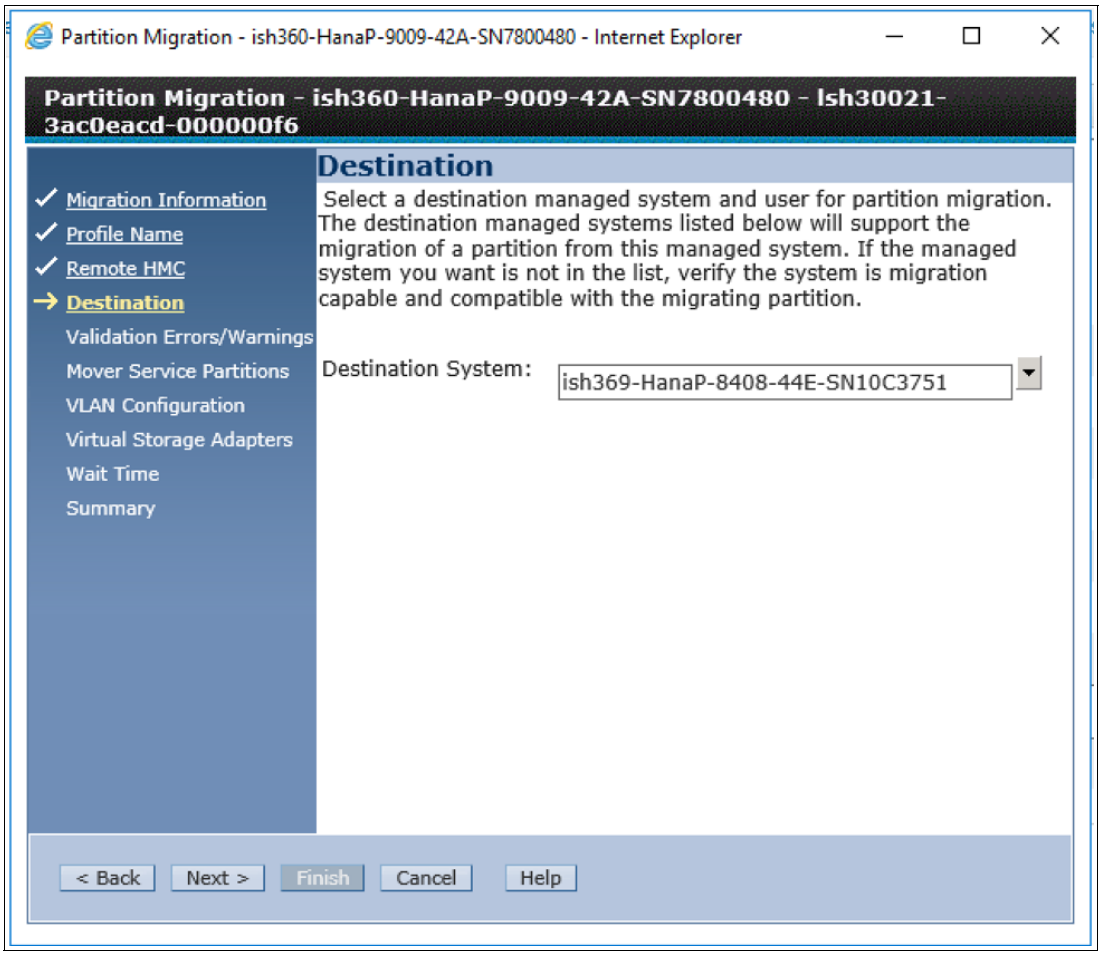


Figure 3-17 Partition Migration: Destination selection

4. The Validation Errors/Warnings step runs through many tests to check whether an LPM can be run successfully, as shown in Figure 3-18. If this step indicates an error, an LPM cannot be performed between the systems. Carefully evaluate the situation and correct any errors.

Note: Checks that produce a warning might or might not inhibit a successful migration. Evaluate the warnings carefully and validate potential issues.

Figure 3-18 Partition Migration: Validation of errors and warnings

Figure 3-18 on page 37 and Figure 3-19 indicate no errors, but show a few warnings. This case does not prohibit a successful migration.

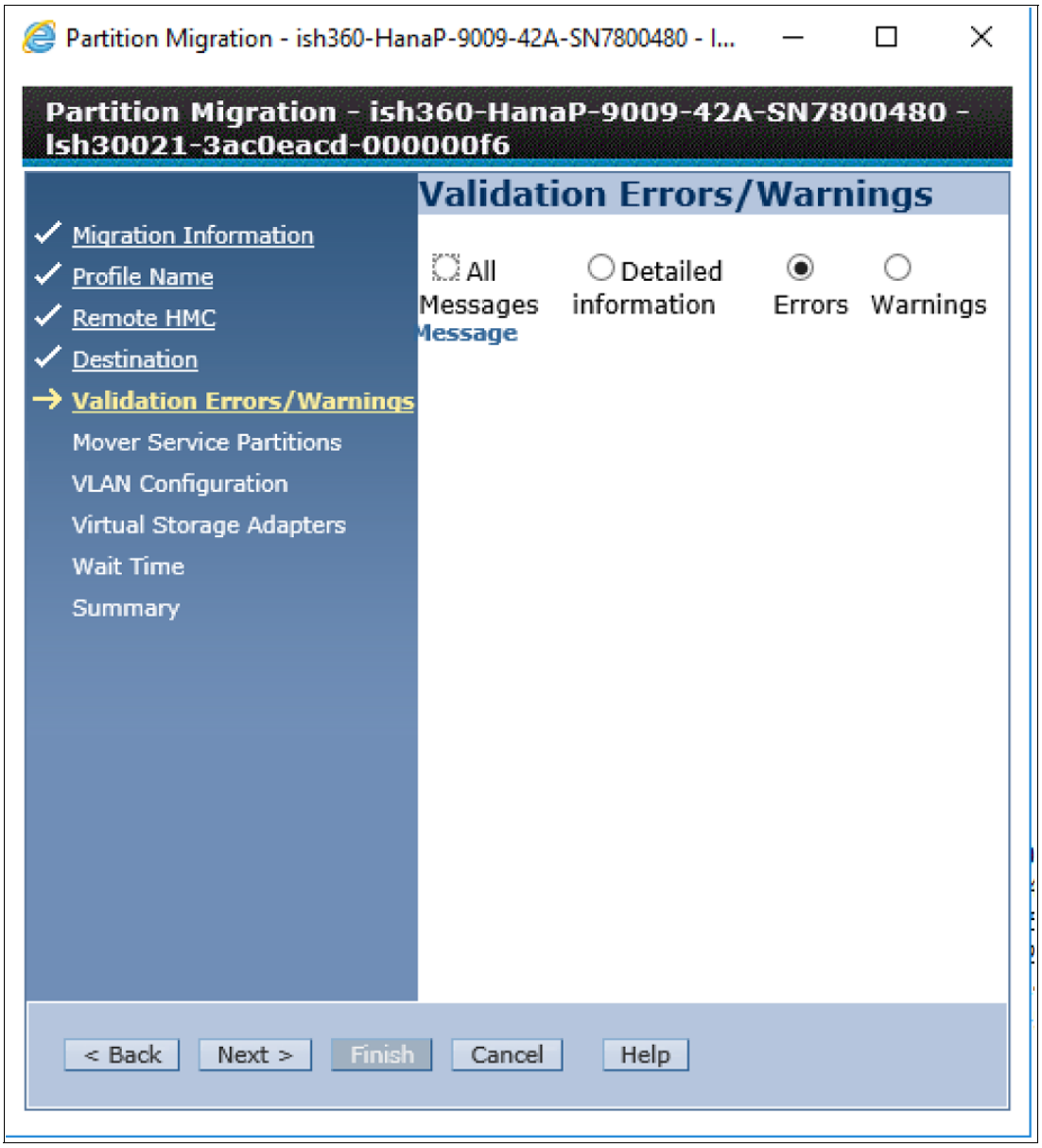


Figure 3-19 Validation Errors/Warnings window: Errors

5. The Mover Service Partitions step enables the definition of MSPs. MSPs are VIOS installations on Power Systems servers that perform the actual work during the migration. One MSP must be selected on the source system and one on the destination system, as shown in Figure 3-20. By default, the HMC preselects one VIOS, but it can be changed in this step.

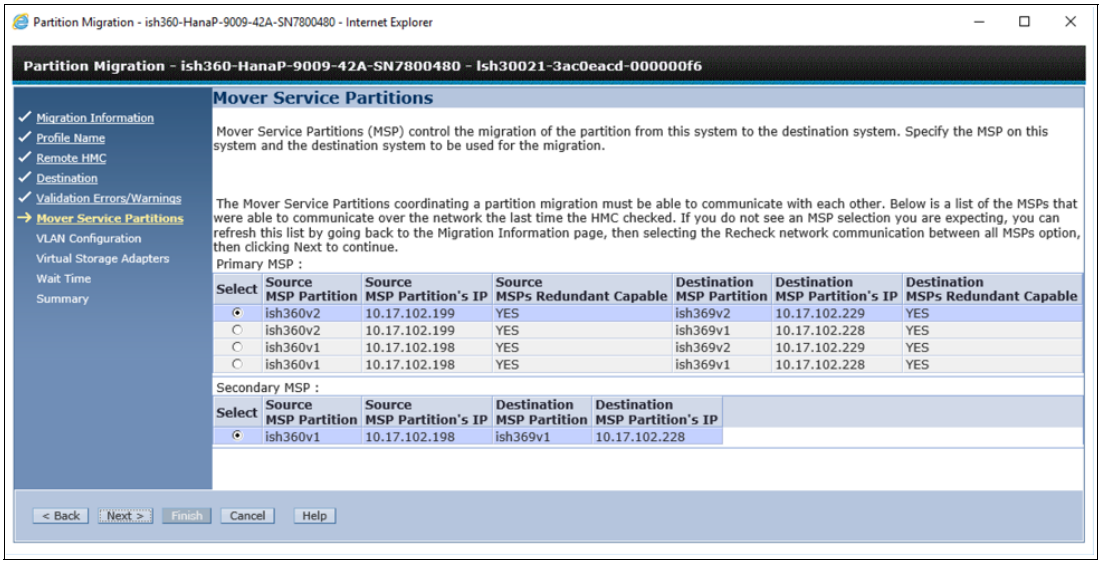


Figure 3-20 Mover Service Partitions window

6. The network information is gathered, as shown in Figure 3-21.

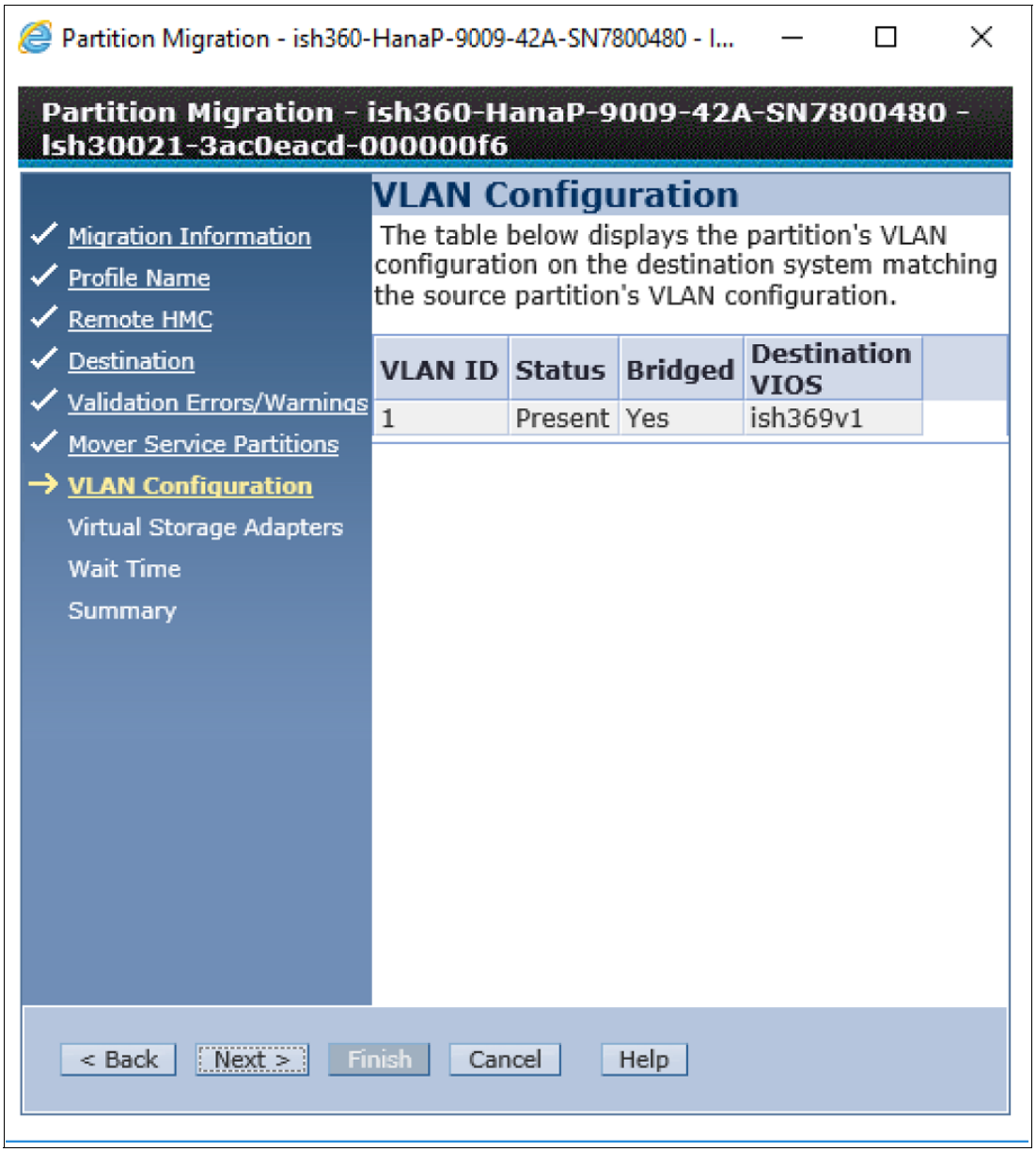


Figure 3-21 VLAN Configuration window: Gathering networking information

7. The storage information is gathered, as shown in Figure 3-22.

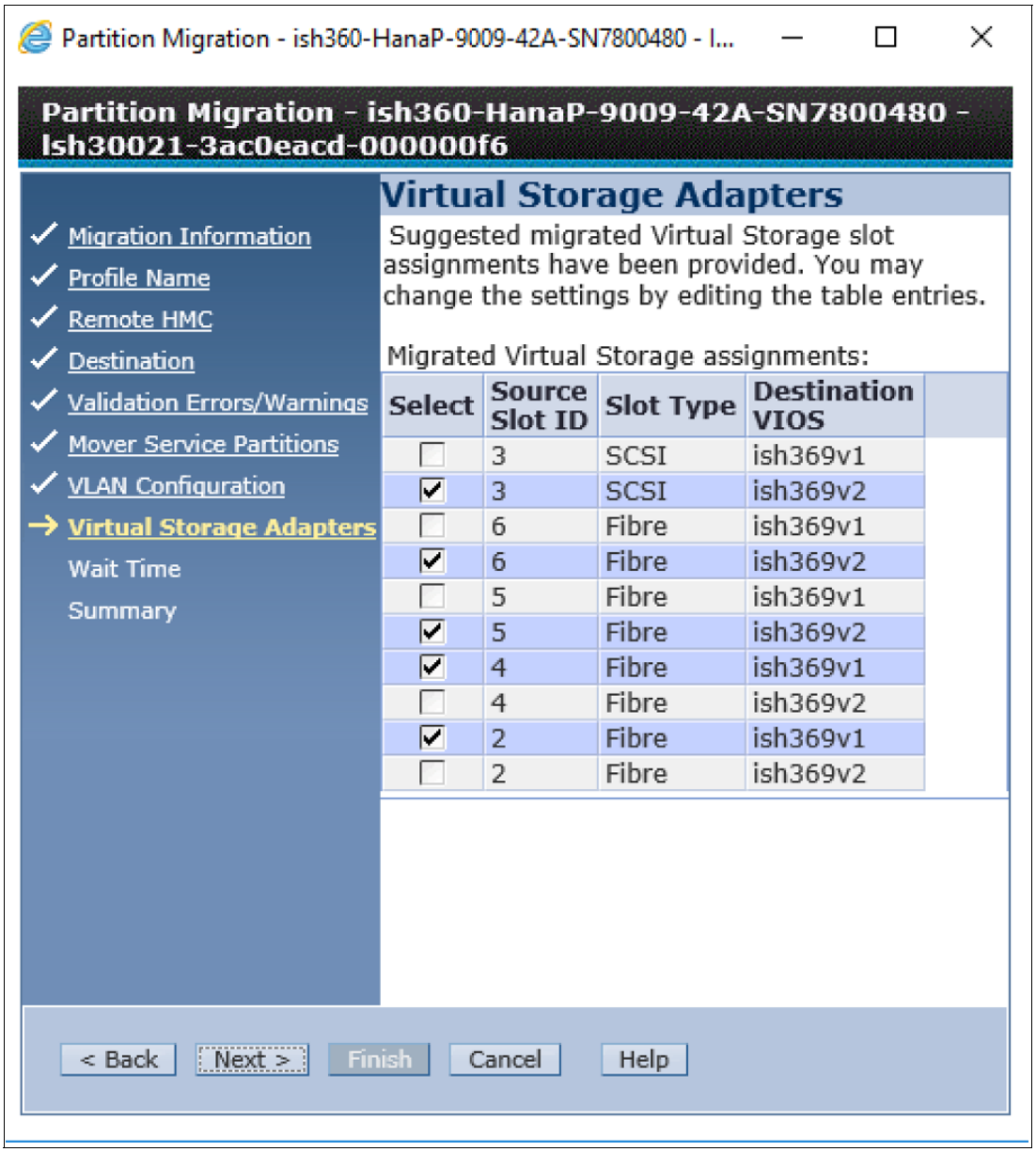


Figure 3-22 Virtual Storage Adapters gathering window

8. In the Wait Time step, a timeout value is set and prefilled by the HMC, as shown in Figure 3-23. The Wait Time value is the time to wait for applications to acknowledge that a migration is happening. This value might have to be increased depending on the configuration of the migrated partition.

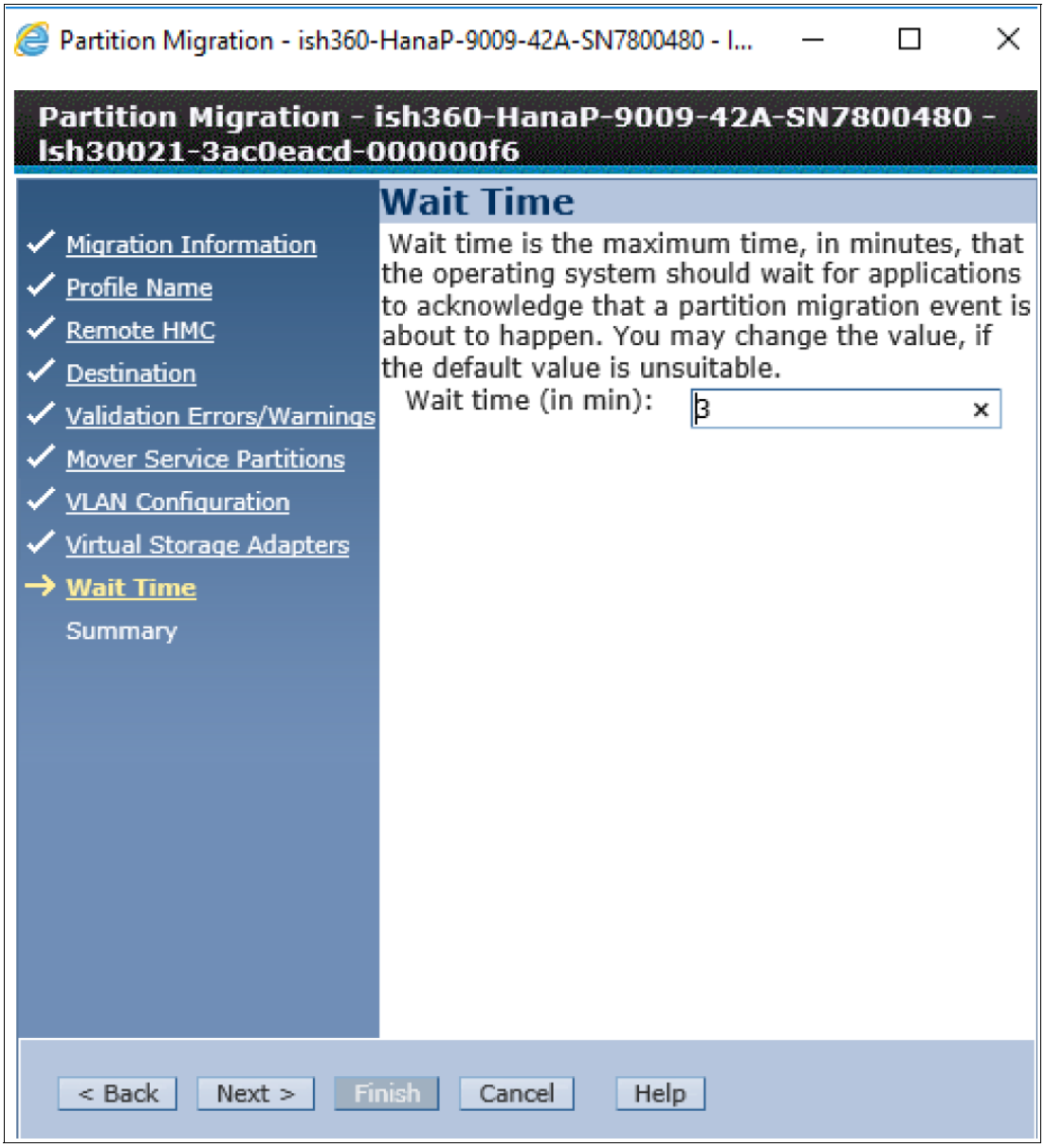


Figure 3-23 Wait Time window: Partition migration timer

9. The Summary step lists all the inputs that are selected during this process, as shown in Figure 3-24. Carefully review the values and click **Finish** to start the migration.

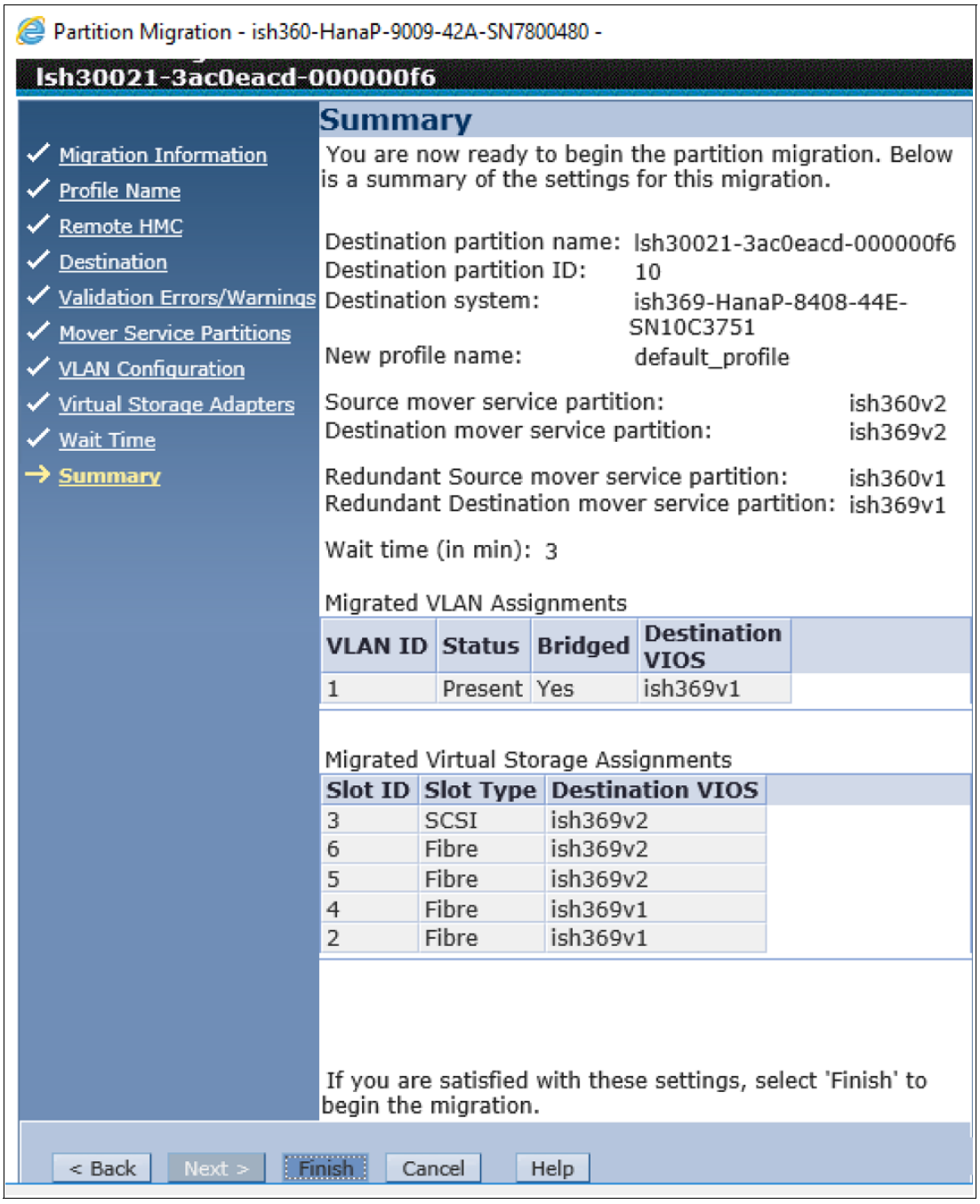


Figure 3-24 Summary window

The migration starts and a progress bar displays, as shown in Figure 3-25.

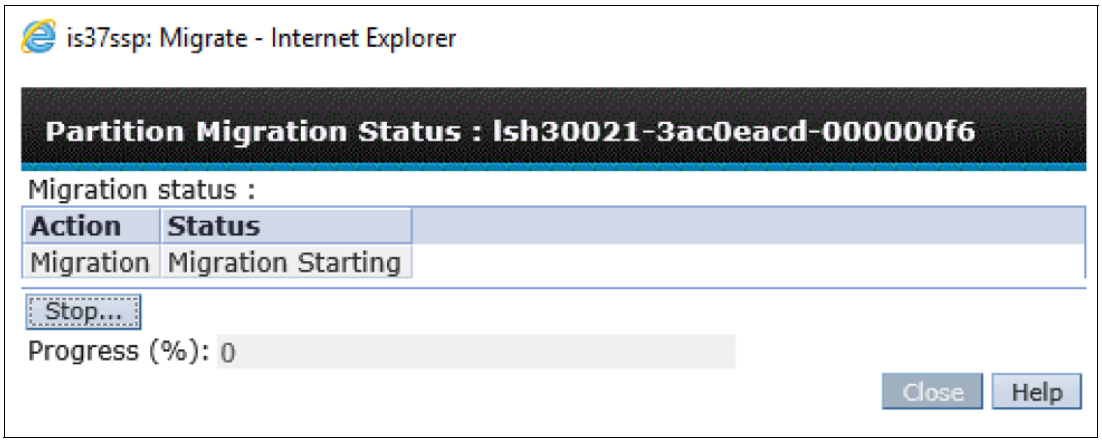


Figure 3-25 Partition Migration Status window

The progress bar continues to count up to 100% until LPM successfully completes. Figure 3-26 shows a successfully completed migration.

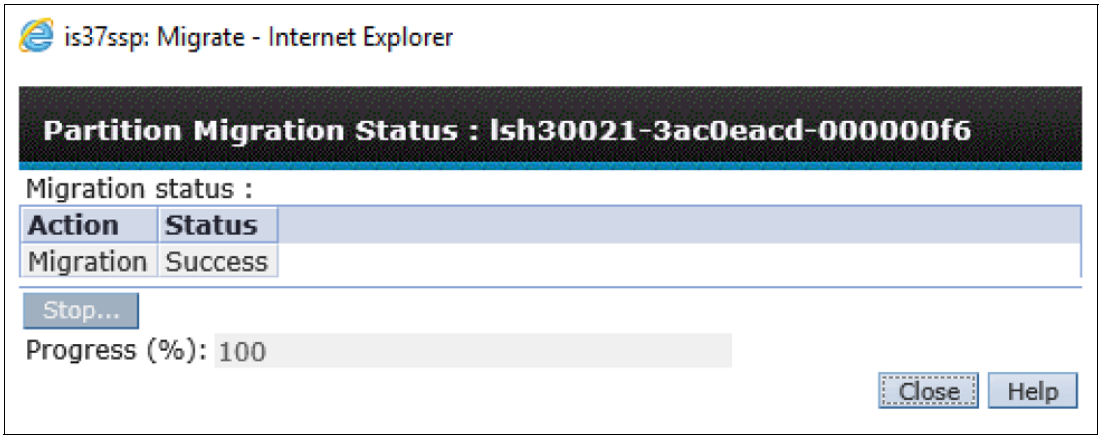


Figure 3-26 Partition Migration Status window: Progress completion

After the successful migration, LPAR lsh30021 is no longer listed on the managed system ish360, as shown in Figure 3-27.

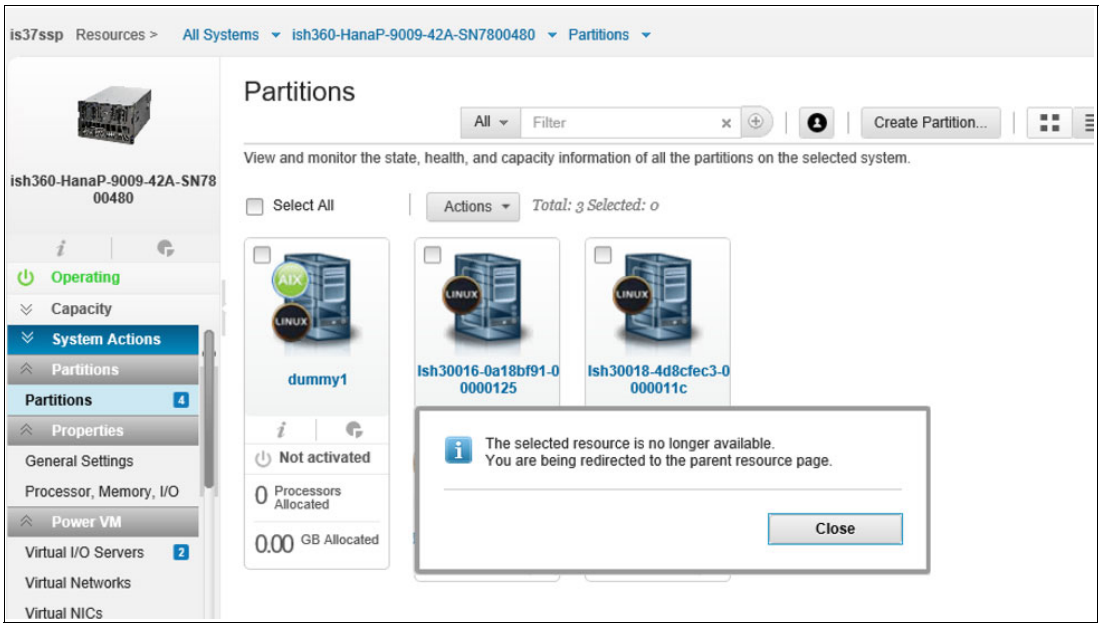


Figure 3-27 Partitions information window

lsh40021 is now available and running on the destination system ish369, as shown in Figure 3-28. The entire migration was performed through a GUI, and the user did not need in-depth knowledge of the technical details of the migration.

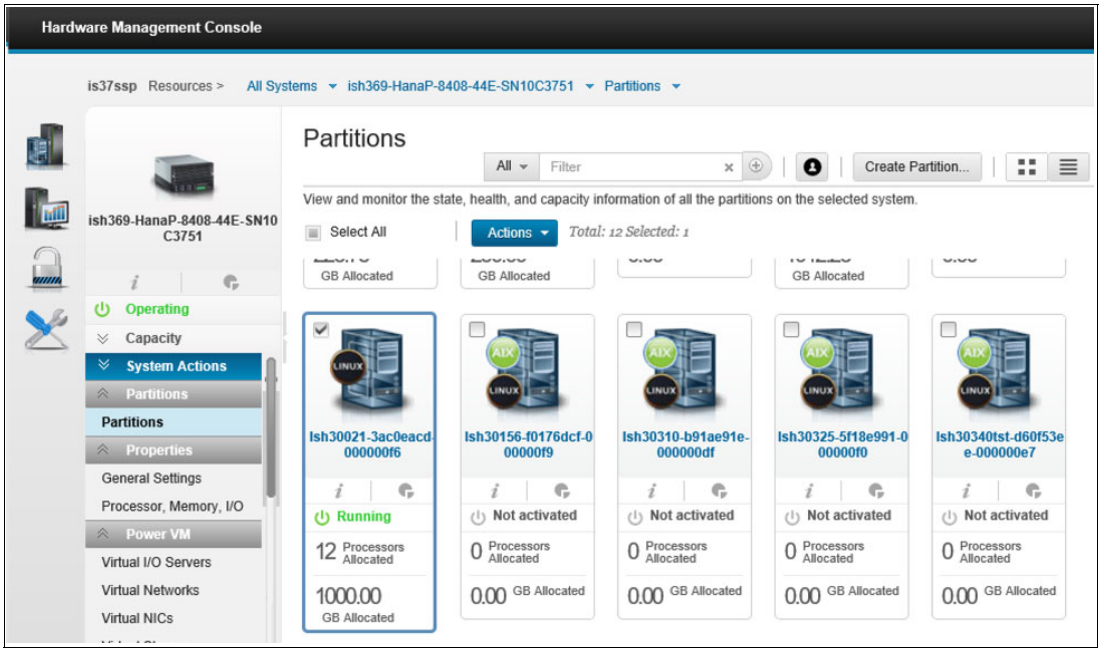


Figure 3-28 Partitions information window: Partition now in destination

10. Log in to the LPAR and perform any functional checks to ensure that the applications are running smoothly on the new host.

    As the partition is likely on different CPUs compared to the old host, the NUMA configuration is likely different, and can be checked again, as shown in Example 3-6.

    Although on the old host all CPUs were allocated on node 1, now 48 logical CPUs are on node 0 and 48 logical CPUs are on node 1. All memory is still on node 1. SAP HANA was stopped before the migration because of this kind of scenario. SAP HANA assumes that CPUs 0 - 48 are still on node 1, and NUMA optimizations within SAP HANA that help performance on the old host might now negatively impact performance because the topology is not as SAP HANA assumes.

Example 3-6 Checking the NUMA configuration

```
[root@lsh30021 ~]# numactl --hardware
available: 2 nodes (0-1)
node 0 cpus: 0 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28
29 30 31 32 33 34 35 36 37 38 39 40 41 42 43 44 45 46 47
node 0 size: 0 MB
node 0 free: 0 MB
node 1 cpus: 48 49 50 51 52 53 54 55 56 57 58 59 60 61 62 63 64 65 66 67 68 69 70 71 72 73
74 75 76 77 78 79 80 81 82 83 84 85 86 87 88 89 90 91 92 93 94 95
node 1 size: 1022928 MB
node 1 free: 794788 MB
node distances:
node   0   1
  0:  10  40
  1:  40  10
[root@lsh30021 ~]#
```

SAP HANA starts successfully in the new topology, as shown in Example 3-7.

Example 3-7 Starting SAP HANA

```
db1adm@lsh30021:/usr/sap/DB1/HDB00> HDB start

StartService
Impromptu CCC initialization by 'rscpCInit'.
  See SAP note 1266393.
OK
OK
Starting instance using: /usr/sap/DB1/SYS/exe/hdb/sapcontrol -prot NI_HTTP -nr 00 -function StartWait 2700
2


24.05.2019 08:43:29
Start
OK

24.05.2019 08:45:33
StartWait
OK
db1adm@lsh30021:/usr/sap/DB1/HDB00> HDB proc
USER       PID  PPID %CPU    VSZ   RSS COMMAND
db1adm    1148 21902  0.0 112064  4160  \_ /bin/sh /usr/sap/DB1/HDB00/HDB proc
db1adm    1181  1148  0.0 112064  1024      \_ /bin/sh /usr/sap/DB1/HDB00/HDB proc
db1adm   31779     1  0.0  10048  4928 sapstart pf=/hana/shared/DB1/profile/DB1_HDB00_lsh30021.sap.corp
db1adm   31801 31779  0.3 412416 78080  \_ /usr/sap/DB1/HDB00/lsh30021.sap.corp/trace/hdb.sapDB1_HDB00 -d
-nw -f /usr/sap/DB1/HDB00/lsh30021.sap.corp/daemon.ini
pf=/usr/sap/DB1/SYS/profile/DB1_HDB00_lsh30021.sap.corp
```

```
db1adm   31819 31801 92.3 8777152 3764992      \_ hdbnameserver
db1adm   32066 31801  2.1 4613056 197632      \_ hdbcompileserver
db1adm   32069 31801  0.9 4849088 202368      \_ hdbpreprocessor
db1adm   32117 31801  719 24041984 20010240      \_ hdbindexserver -port 30003
db1adm   32120 31801 10.4 7213504 1348224      \_ hdbxsengine -port 30007
db1adm     507 31801 25.3 6080704 627520      \_ hdbwebdispatcher
db1adm   18094     1  0.0 6829120 46528 hdbrsutil  --start --port 30003 --volume 3 --volumesuffix
mnt00001/hdb00003.00003 --identifier 1573188216
db1adm   17518     1  0.0 668672 40448 hdbrsutil  --start --port 30001 --volume 1 --volumesuffix
mnt00001/hdb00001 --identifier 1573188180
db1adm   17139     1  0.0 610560 39936 /usr/sap/DB1/HDB00/exe/sapstartsrv
pf=/hana/shared/DB1/profile/DB1_HDB00_lsh30021.sap.corp -D -u db1adm
db1adm@lsh30021:/usr/sap/DB1/HDB00>
```

Example 3-8 shows SAP HANA functioning in the new partition.

Example 3-8 Checking the SAP HANA functions

```
db1adm@lsh30019:/usr/sap/DB1/HDB00>  sapcontrol -nr 00 -function GetProcessList

24.05.2019 09:18:34
GetProcessList
OK
name, description, dispstatus, textstatus, starttime, elapsedtime, pid
hdbdaemon, HDB Daemon, GREEN, Running, 2019 05 24 08:43:30, 0:35:04, 31801
hdbcompileserver, HDB Compileserver, GREEN, Running, 2019 05 24 08:43:39, 0:34:55, 32066
hdbindexserver, HDB Indexserver-QH1, GREEN, Running, 2019 05 24 08:43:40, 0:34:54, 32117
hdbnameserver, HDB Nameserver, GREEN, Running, 2019 05 24 08:43:31, 0:35:03, 31819
hdbpreprocessor, HDB Preprocessor, GREEN, Running, 2019 05 24 08:43:39, 0:34:55, 32069
hdbwebdispatcher, HDB Web Dispatcher, GREEN, Running, 2019 05 24 08:45:01, 0:33:33, 507
hdbxsengine, HDB XSEngine-QH1, GREEN, Running, 2019 05 24 08:43:40, 0:34:54, 32120
db1adm@lsh35263le.wdf:/usr/sap/DB1/HDB00>
```

The LPM process works much the same way if an *offline* migration is performed. The only difference is that the OS in the partition is not active during the migration.

To provide a greater level of flexibility for Power Systems server owners, it is possible to migrate up to 32 LPARs in parallel.

The latest version of VIOS (Version 3.1) introduces data encryption and compression support during migration, and 100 Gb Ethernet cards. These items increase data security and reduce the migration duration. Starting with server firmware FW920, HMC V9 R1.920.0 and Novalink 1.0.0.10, the data that is sent between the servers is automatically encrypted for better security and compressed for better performance.

Fast data transfer speeds are especially important in SAP HANA environments because the in-memory character of SAP HANA allocates large amounts of memory that must be transferred.

For more information about these features, see Live Partition Mobility (LPM) Data Encryption and Compression.

The following links provide important SAP Notes for LPM:

- SAP Note 1102760
- SAP Note 994025

3.4 Dynamic Platform Optimizer and dynamic logical partition operations

This section describes features that enhance LPARs and provide more capabilities for applications running on Power Systems servers.

3.4.1 Viewing the current topology information

From a Linux LPAR, the CPU and memory resource topology that is detected by the OS can be inspected by running the **numactl** command, as shown in Example 3-9.

Example 3-9 Viewing the hardware topology information

```
# numactl –hardware
available: 2 nodes (0-1)
node 0 cpus: 0 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25
26 27 28 29 30 31
node 0 size: 0 MB
node 0 free: 0 MB
node 1 cpus: 32 33 34 35 36 37 38 39 40 41 42 43 44 45 46 47 48 49 50 51 52 53 54
55 56 57 58 59 60 61 62 63
node 1 size: 260811 MB
node 1 free: 141305 MB
node distances:
node   0   1
  0:  10  20
  1:  20  10
```

This **numactl** output (Example 3-9) shows an LPAR with CPU resources that are equally allocated on NUMA node 0 and NUMA node 1. Memory resources are allocated only from NUMA node 1.

A system-level view of the topology can be collected from a system resource dump that is performed by using the HMC, as shown in Example 3-10.

Example 3-10 System topology view that is collected from a system resource dump

```
hscroot@hmc1:~>  startdump -m SystemA-9119-MME-SN65C4D3D  -t resource -r 'hvlpconfigdata -affinity -domain'
hscroot@hmc1:~> cat /dump/RSCDUMP.65C4D3D.04000004.20190527045853

|-----------|----------------------|---------------|------|---------------|---------------|-------|
|   Domain  |      Procs       Units |     Memory    |      |   Proc Units  |     Memory    | Ratio |
| SEC | PRI | Total | Free  | Free  | Total |  Free | LP   |  Tgt  |  Aloc | Tgt   |  Aloc |       |
|-----|-----|-------|-------|-------|-------|-------|------|-------|-------|-------|-------|-------|
|   0 |     |  4000 |   400 |     0 |  8192 |  1452 |      |       |       |       |       |  1772 |
|     |   0 |  1000 |   200 |     0 |  2048 |   834 |      |       |       |       |       |  2036 |
|     |     |       |       |       |       |       |    2 |   200 |   200 |    16 |    16 |       |
|     |     |       |       |       |       |       |    6 |   400 |   400 |       |       |       |
|     |     |       |       |       |       |       |    7 |   100 |   100 |   640 |   640 |       |
|     |     |       |       |       |       |       |   13 |   100 |   100 |   512 |   512 |       |
|     |     |       |       |       |       |       |   15 |       |       |     1 |     1 |       |
|     |   1 |  1000 |   200 |     0 |  2048 |   452 |      |       |       |       |       |  1103 |
|     |     |       |       |       |       |       |    1 |   200 |   200 |    16 |    16 |       |
|     |     |       |       |       |       |       |    6 |   400 |   400 |  1024 |  1024 |       |
|     |     |       |       |       |       |       |  199 |   200 |   200 |   512 |   512 |       |
|     |   2 |  1000 |     0 |     0 |  2048 |   154 |      |       |       |       |       |     0 |
```

```
|       |       |       |       |       |       |       |       |     4 |   200 |   200 |   160 |   160 |       |
|       |       |       |       |       |       |       |       |     5 |   400 |   400 |  1024 |  1024 |       |
|       |       |       |       |       |       |       |       |    21 |   400 |   400 |   640 |   640 |       |
|       |     3 |  1000 |     0 |     0 |  2048 |    12 |       |       |       |       |       |       |     0 |
|       |       |       |       |       |       |       |       |     3 |   300 |   300 |  1024 |  1024 |       |
|       |       |       |       |       |       |       |       |    15 |   100 |   100 |   206 |   206 |       |
|       |       |       |       |       |       |       |       |    16 |   400 |   400 |    40 |    40 |       |
|       |       |       |       |       |       |       |       |    19 |   100 |   100 |   640 |   640 |       |
|       |       |       |       |       |       |       |       |    70 |   100 |   100 |    16 |    16 |       |
|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|
|     1 |       |  4000 |  1000 |     0 |  8192 |  3184 |       |       |       |       |       |       |  1554 |
|       |     4 |  1000 |   100 |     0 |  2048 |  1134 |       |       |       |       |       |       |  5537 |
|       |       |       |       |       |       |       |       |     9 |   400 |   400 |    16 |    16 |       |
|       |       |       |       |       |       |       |       |    15 |   100 |   100 |   201 |   201 |       |
|       |       |       |       |       |       |       |       |    20 |   400 |   400 |   640 |   640 |       |
|       |     5 |  1000 |   500 |     0 |  2048 |  1204 |       |       |       |       |       |       |  1175 |
|       |       |       |       |       |       |       |       |    10 |   100 |   100 |    64 |    64 |       |
|       |       |       |       |       |       |       |       |    15 |   100 |   100 |   208 |   208 |       |
|       |       |       |       |       |       |       |       |    26 |   300 |   300 |   480 |   480 |       |
|       |     6 |  1000 |   300 |     0 |  2048 |   332 |       |       |       |       |       |       |   540 |
|       |       |       |       |       |       |       |       |     8 |   400 |   400 |   640 |   640 |       |
|       |       |       |       |       |       |       |       |    15 |   100 |   100 |   206 |   206 |       |
|       |       |       |       |       |       |       |       |    22 |   100 |   100 |   640 |   640 |       |
|       |       |       |       |       |       |       |       |    26 |   100 |   100 |   160 |   160 |       |
|       |     7 |  1000 |   100 |     0 |  2048 |   514 |       |       |       |       |       |       |  2509 |
|       |       |       |       |       |       |       |       |    15 |   100 |   100 |   202 |   202 |       |
|       |       |       |       |       |       |       |       |    17 |   400 |   400 |   640 |   640 |       |
|       |       |       |       |       |       |       |       |    24 |   400 |   400 |   640 |   640 |       |
|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|
```

Example 3-10 on page 48 shows resource dump information for LPAR 6, which matches the information that is collected by running **numact** (Example 3-9 on page 48).

3.4.2 Dynamic logical partition

This section describes how to identify partitions that are enabled for DLAPR and its capabilities.

DLPAR-capable LPARs

Example 3-11 shows the **lspartition -dlpar** command that is run in the HMC, indicating whether an LPAR can run DLPAR operations.

Example 3-11 Checking the LPAR capabilities for dynamic logical partitioning

```
# lspartition -dlpar
<#178> Partition:<1*9179-MHD*10226D4, ish356v1.wdf.sap.corp, 10.17.102.142>
       Active:<1>, OS:<AIX, 6.1, 6100-09-06-1543>, DCaps:<0x14f8f>, CmdCaps:<0x4000003b,
0x3b>, PinnedMem:<2111>
<#179> Partition:<3*9080-M9S*7816D78, lsh30220.wdf.sap.corp, 10.17.103.26>
       Active:<1>, OS:<Linux/SuSE, 4.4.140-94.42-default, 12>, DCaps:<0x0>, CmdCaps:<0x0,
0x0>, PinnedMem:<0>
<#180> Partition:<20*9119-MME*65C4D3D, lsh30078.wdf.sap.corp, 10.17.102.144>
       Active:<1>, OS:<Linux/SuSE, 4.4.162-94.72-default, 12>, DCaps:<0x2c7f>,
CmdCaps:<0x40000019, 0x19>, PinnedMem:<0>
<#181> Partition:<2*9119-MHE*1092627, ish329v2.wdf.sap.corp, 10.17.102.21>
       Active:<1>, OS:<AIX, 6.1, 6100-09-06-1543>, DCaps:<0x14f8f>, CmdCaps:<0x4000003b,
0x3b>, PinnedMem:<2507>
```

A **Dcaps** value of 0x0 indicates that DLPAR operations are not available for the LPAR. The most common cause for not supporting DLPAR operations is an inactive Resource Monitoring and Control (RMC) connection. To enable this connection, install the DynamicRM and ServiceRM file sets in the Linux OS.

The following RPMs must be installed in the order listed:

- src
- rsct.core.utils
- rsct.core
- csm.core
- csm.cleint
- devices.chrp.base.ServiceRM
- DynamicRM

DLPAR operation examples

The **chhwres** command on the HMC is used to perform DLPAR operations. For a complete description of the command, run **man chhwres**.

Use the following syntax to add, remove, or move memory:

```
chhwres -r mem -m managed-system -o {a | r | m} {-p partition-name | --id
partition-ID} [{-t target-partition-name | --tid target-partition-ID}] -q quantity
[-w wait-time] [-d detail-level] [--force]
```

For example, the following command removes 256 GB memory from the LPAR with ID 6 on the managed system:

```
chhwres -r mem -m SystemA-9119-MME-SN65C4D3D -o r –id 6 –q 256
```

Use the following syntax to add, remove, or move processing resources:

```
chhwres -r proc -m managed-system -o {a | r | m} {-p partition-name | --id
partition-ID} [{-t target-partition-name | --tid target-partition-ID}]
[--procs quantity] [--procunits quantity] [--5250cpwpercent percentage]
[-w wait-time] [-d detail-level] [--force]
```

For example, the following command adds three processors to the LPAR with ID 6 on the managed system:

```
chhwres -r proc -m SystemA-9119-MME-SN65C4D3D -o a –id 6 –procs 3
```

3.4.3 Dynamic Platform Optimizer

This section describes Dynamic Platform Optimizer (DPO) and its characteristics.

Affinity scores

The command **lsmemopt** is used on the HMC to determine affinity scores for the system as a whole or for individual LPARs: It has the following syntax:

```
lsmemopt -m managed-system [-r {sys | lpar}] [-o {currscore | calcscore}]
[-p partition-names | --id partition-IDs] [-x partition-names | --xid
partition-IDs]
[--filter "filter-data"]  [-F [attribute-names] [--header]] [--help]
```

For a complete description of this command, run **man lsmemopt**.

The affinity score is a number 0 - 100, where 0 represents the worst affinity and 100 represents perfect affinity.

When used with the option **currscore**, the current processor to memory affinity scores are shown.

When used with the option **calcscore**, the predicted or potential affinity score is calculated and displayed, which is an estimate of the affinity score that can be achieved after a DPO operation.

DPO priority order

DPO operations are run in order of priority. To determine the priority, the LPAR attribute **affinity_group_id** is evaluated. This attribute supports values 1 (lowest priority) - 255 (highest priority).

Multiple LPARs can share a value, which directs the hypervisor to attempt to colocate the LPARs on the same chip or node as much as possible. Among LPARs with the same **affinity_group_id**, priority is set based on CPU and memory resource allotments.

Starting DPO

The HMC command **optmem** runs DPO operations. It has the following syntax:

```
optmem -m managed-system -o start -t affinity [-p partition-names | --id
partition-IDs] [-x partition-names | --xid partition-IDs]
```

Where:

- **-x partition-names** or **--xid partition-IDs** specifies the list of LPAR names or LPAR IDs that must *not* be affected by the optimization operation.
- **-p partition-names** or **--id partition-IDs** specifies the list of LPAR names or LPAR IDs that are selected for optimization.

Viewing DPO progress

Depending on the size and usage of LPARs, DPO operations can take a significant amount of time. The status can be shown by running the **lsmemopt** command, as shown in Example 3-12.

Example 3-12 Showing the status of DPO

```
# lsmemopt –m SystemA-9119-MME-SN65C4D3D
in_progress=0,status=Finished,type=affinity,opt_id=1,
progress=44,requested_lpar_ids=none,protected_lpar_ids=none,"impacted_lpar_ids=6"
```

4

Kernel-based virtual machine and Red Hat Virtualization management and operations

The Kernel-based virtual machine (KVM) and Red Hat Virtualization (RHV) features provide fast virtualization for Linux guests on small- and medium-sized systems.

This chapter contains information about the bare metal firmware configuration and bootstrap installation of the KVM host operating system (OS) and the RHV software.

This chapter contains the following topics:

- Bare metal installation: Getting started and preparing the server
- Loading and initializing Linux on bare metal Power Systems hardware
- Installing RHEL OS as a KVM host on bare metal Power Systems hardware
- Red Hat Virtualization

4.1 Bare metal installation: Getting started and preparing the server

The bare metal installation (BMI) is a direct installation of an OS on top of the firmware. In this example, you are installing the Red Hat Enterprise Linux (RHEL) OS on local disks.

This section contains information about configuring the installation, firmware updates of the server, and integrated RAID adapters.

Note: The system firmware is a combination of the baseboard management controller (BMC) firmware and the processor NOR (PNOR) firmware. To update the system firmware, update both the BMC firmware and the PNOR firmware by using the BMC.

4.1.1 Preparing the bare metal firmware

To prepare the bare metal firmware, see the “Linux information for IBM systems” at IBM Knowledge Center, which helps you to install Linux on a Power Systems server (Figure 4-1).

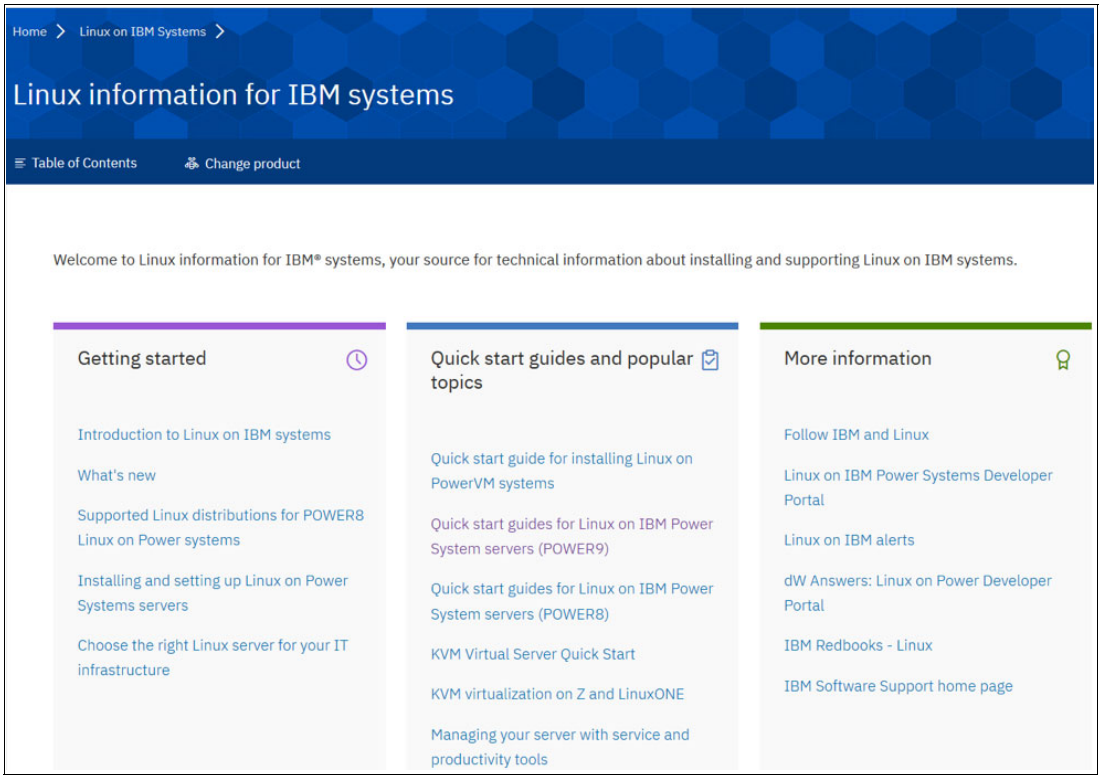


Figure 4-1 Linux information for IBM Systems page

4.1.2 Petitboot bootloader

Petitboot is a bootloader that is based on kexec. It is used in IBM POWER8 and POWER9 processor-based servers to boot the initial Linux OS.

After the Power Systems server starts, the Petitboot bootloader scans local boot devices and network interfaces to find boot options that are available to the system. Petitboot presents a list of boot options for the running system. In the case of a static IP or no provided boot arguments for a network boot server, you must provide the details. To configure Petitboot to find the boot device, see IBM Knowledge Center.

The Petitboot configuration can be adjusted. For example, to change the amount of time before Petitboot automatically boots, see IBM Knowledge Center.

After deciding to boot the ISO media for the Linux distribution of your choice, the installer wizard for this Linux distribution opens and is used to set up disk options, the root password, time zones, and other items.

For more information about the Petitboot bootloader program, see Petitboot: A kexec based bootloader.

4.1.3 Configuring the baseboard management controller IP address

To set up or enable a network connection to the BMC[1], use the Petitboot bootloader interface and complete the following steps.

Important: This address is not the IP address of the Petitboot OS. For more information, see 4.1.10, “Supermicro RAID Controller configuration and firmware update” on page 65.

1. Power on your server, and the Petitboot bootloader menu opens. This process takes about 1 - 2 minutes to complete. Do not walk away from your system. When Petitboot loads, push any key to interrupt the boot process.
2. At the Petitboot bootloader main menu, select `Exit to shell`.
3. Run the **`ipmitool lan print 1`** command. If this command returns an IP address, verify that it is correct and continue. To set a static IP address, complete these steps:
   a. Set the mode to static by running this command:

      ```
      ipmitool lan set 1 ipsrc static
      ```

   b. Set your IP address by running this command:

      ```
      ipmitool lan set 1 ipaddr <ip_address>
      ```

      Where **`ip_address`** is the static IP address that you are assigning to this system.

   c. Set your netmask by running this command:

      ```
      ipmitool lan set 1 netmask <netmask_address>
      ```

      Where **`netmask_address`** is the netmask for the system.

[1] Source: ftp://public.dhe.ibm.com/systems/power/docs/hw/p9/p9eih.pdf

d. Set your gateway server by running this command:

```
ipmitool lan set 1 defgw ipaddr <gateway_server>
```

Where **gateway_server** is the gateway for this system.

e. Confirm and check the IP address by running this command:

```
ipmitool lan print 1
```

Note: This network interface is not active until after you perform the following steps.

4. To reset your firmware, run the following command:

```
ipmitool mc reset cold
```

This command must complete before you continue the process, but it does not return any information. To verify that this command completes, ping your system BMC address (the same IP address that is used in your **ipmitool** command).

If your ping does not return successfully within a reasonable amount of time (2 - 3 minutes), try these additional steps:

a. Power off your system by running this command:

```
ipmitool power off
```

b. Unplug the power cords from the back of the system. Wait 30 seconds, plug them back in, and start the system to start the BMC.

Logging on to the baseboard management controller GUI

To log on to the BMC GUI, complete the following steps:

1. Open a supported web browser. In the address bar, enter the host name or the IP address of the BMC to which you want to connect. For example, use the format *http://1.2.3.4* or *http://hostname.example.com* in the address bar of the web browser.
2. On the BMC logon page, enter the user name and password that is assigned to you:
   - The default user ID is ADMIN and the default password is ADMIN (admin).
   - If you forgot your assigned password, enter your user name and click **Forgot Password**? Follow the instructions.
3. Click **Login**.

4.1.4 Setting up password controls for the BMC

Password controls can be set up for BMC LAN access on Power Systems servers. This section shows how to set up password control for two users (the default user with user ID 1 and the null user) in the LAN channel.

Note: To reduce vulnerability, the Intelligent Platform Management Interface (IPMI) LAN interface must be enabled only in a trusted environment where the system is secure or where it is connected to a dedicated secure or private network.

The BMC can be configured to support multiple users and passwords for all channels *except* for the Open channel. Typically, the same user and same password can be used for all the BMC channels. Instructions to set up password control for other channels are not included in this example. The instructions can be used only for the LAN channel.

User IDs and privilege levels are unique for each channel. To view the current user IDs that are in use and the related information for the LAN channel (0x1), run the following command:

```
# ipmitool user list 1
ID  Name      Callin Link Auth IPMI Msg  Channel Priv Limit
1   USERID     true  false   true    ADMINISTRATOR
```

Note: On all IBM BMCs, the default user ID 1 is USERID with a password of PASSW0RD.

To change the name of user ID 1, run the following command:

```
# ipmitool user set name 1 <New User ID>
```

To set a new password for user ID 1, run the following command:

```
# ipmitool user set password 1 ipmitool user set password 1 <New Password>
```

You can also use a null user for anonymous login. To change the password for the null user (user ID 1) on the LAN channel, run the following command:

```
# ipmitool lan set 1 password <New Password>
```

You can list the users that you set up and find the new name (user ID) for the user ID 1 user. The null user is not listed by running the following command when it is disabled in the BMC BIOS settings:

```
# ipmitool user list 1
```

4.1.5 Updating firmware by using ipmitool

The IPMI is an open standard for monitoring, logging, recovery, inventory, and control of hardware that is implemented independently of the main CPU, BIOS, and OS. The LC Power Systems servers provide one 10M/100M baseT IPMI port.

The **ipmitool** is a utility for managing and configuring devices that support IPMI. It provides a simple command-line interface (CLI) to the service processor. The **ipmitool** can be installed from the Linux distribution packages on your workstation, SourceForge, or another server (preferably on the same network as the installed server).

There is a minimum **ipmitool** code level that is required by the system firmware to manage the system. OpenPOWER requires **ipmitool** V1.8.15 or later to run correctly on the OP910 firmware. The **ipmitool** must be capable of establishing an IPMI V2 session with the IPMI support on the BMC. Verify your **ipmitool** level on your Linux workstation by running the following command:

```
# ipmitool -V
ipmitool version 1.8.18
```

Important: The syntax can be different under Linux and Windows, and connection problems can be caused by outdated firmware.

To connect to your host system with IPMI, you need to know the IP address of the server and have a valid password. To power on the server with **ipmitool**, complete these steps:

1. Open a terminal program.
2. Power on your server by running **ipmitool**:

   ```
   ipmitool -I lanplus -H bmc_ip_address -U ipmi_userid -P ipmi_password power on
   ```

Tip: If there are connection problems, it might be due to outdated firmware.

3. Activate your IPMI console:

```
ipmitool -I lanplus -H bmc_ip_address  -U ipmi_userid -P ipmi_password sol
activate
```

For more information about the **ipmitool** and the firmware update process, see IBM Knowledge Center.

Here is an example of updating the firmware by using the **ipmitool**:

1. Verify your current firmware level by running the following command:

```
ipmitool -H bmc_ip_ipaddress -I lan -U ipmi_user -P ipmi_password fru list
```

   Look for the section titled `FRU Device Description: System Firmware (ID xx)`. Then, find the value under `Product version`. This value is your firmware level.

2. Download your firmware fix from Fix Central. Click **Select Product**. Click **Power** from **Product Group** and **Scale-out LC** from the **Product**. Select your server model and the fixes that you want to download. Click **Continue**.

3. Your system must be in standby state to update the firmware. Power off the system by running the following command:

```
ipmitool -H bmc_ip_ipaddress -I lan -U ipmi_user-P ipmi_password chassis power
off
```

4. Reset the hardware to a known state by running the following command:

```
ipmitool -H bmc_ip_ipaddress -I lan -U ipmi_user -P ipmi_password command mc
reset cold
```

   This command must complete before continuing the process. To verify that this command completed, ping your system BMC address (the same IP address that is used in your **ipmitool** command). When the ping returns successfully, continue.

5. To protect the BMC network and IPMI settings, run the following command:

```
ipmitool -H bmc_ip_ipaddress -I lan -U ipmi_user -P ipmi_password raw 0x32 0xba
0x18 0x00
```

   Note: This command does not return any output.

6. To flash the BMC and PNOR firmware, run this command:

```
ipmitool -H bmc_ip_ipaddress -I lan -U ipmi_user -P ipmi_password hpm upgrade
firmware _update_name -z 30000 force
```

   Note: The update process for this firmware version can require up to 30 minutes to complete per BMC.

7. Restart your system by running the following command:

```
ipmitool -H bmc_ip_ipaddress -I lan -U ipmi_user -P ipmi_password chassis power
on
```

Important: If the update process ends before it completes, wait at least 30 minutes before retrying. You can also reset the firmware and try again immediately by running this command:

```
ipmitool mc reset cold
```

Helpful ipmitool commands

The following **ipmitool** commands are useful:

- CLI help:

  ```
  # ipmitool help
  ```

- To display the history of your system event log, run this command:

  ```
  # ipmitool sel list
  ```

- To obtain information and the status of a system locally or remotely, run the **ipmitool chassis** command:

  ```
  # ipmitool chassis status
  ```

4.1.6 Updating the firmware by using the OpenBMC CLI

The process of updating firmware on the OpenBMC managed servers is documented in this section. The following events must happen in sequence:

1. Power off the host.
2. Update and activate the BMC.
3. Update and activate the processor NOR (PNOR) (PNOR flash chip).
4. Restart the BMC (applies a new BMC image).
5. Power on the host (applies a new PNOR image).

The OpenBMC firmware updates (BMC and PNOR) for the servers can be managed by using the **openbmctool** CLI tool. To obtain the **openbmctool** tool, see Scale-out LC System Event Log Collection Tool and complete the following steps:

1. In the search field, enter your machine type and model. Then, click the correct product support entry for your system.
2. From the downloads list, click **openbmctool** for your machine type and model.
3. Follow the instructions to install and run the **openbmctool**. You must provide the file locations for the BMC firmware image file and PNOR firmware image file that must be downloaded from Fix Central for the update level that you need.

For more information about the **openbmctool** and the firmware update process, see IBM Knowledge Center.

Here are the commands to extract and run the **openbmctool** installation:

```
# cd firmware-update-directory

# openbmctool -U root -P OpenBmc -H powerbmc firmware flash bmc -f
obmc-witherspoon-ibm-v2.0.ubi.mtd.tar

# openbmctool -U root -P OpenBmc -H powerbmc fru print | grep Activ

# openbmctool -U root -P OpenBmc -H powerbmc firmware flash pnor -f
witherspoon-IBM-OP9_v1.19_1.192.pnor.squashfs.tar
```

```
# openbmctool -U root -P OpenBmc -H powerbmc fru print | grep Activ

# openbmctool -U root -P OpenBmc -H powerbmc mc reset cold
```

Helpful openbmctool commands

Here are some helpful **openbmctool** commands:

- **# openbmctool -H <hostname> -U root -P OpenBmc chassis power -h**
- **# openbmctool -H <hostname> -U root -P OpenBmc chassis power status**
- **# openbmctool -H <hostname> -U root -P OpenBmc chassis power softoff**
- **# openbmctool -H <hostname> -U root -P OpenBmc chassis power on**

4.1.7 Updating the system firmware by using the BMC Web GUI

Another method to update the system firmware is by using the BMC Web GUI. For more information, see IBM Power Systems Scale-out LC Server Firmware.

Note: The system firmware update from the BMC Web GUI is supported on only Google Chrome and Mozilla Firefox browsers.

To update the BMC firmware, complete the following steps:

1. Log in to the BMC by entering the user name and password, and then press Enter.
2. From the Maintenance list on the BMC dashboard, select **BMC Update**, as shown in Figure 4-2.

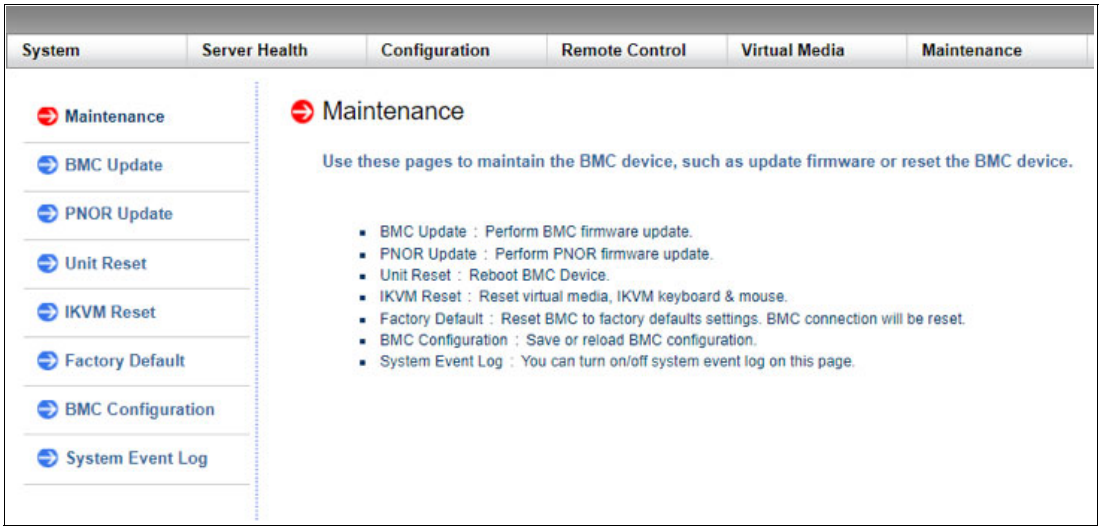


Figure 4-2 BMC Update window

3. In the BMC Update window, select **Enter Update Mode**. Click **OK**.
4. In the BMC Upload window, choose the `.bin` file from your local system folder and click **Upload Firmware**. Wait for the file to be uploaded, and then click **OK**.
5. The existing and new versions of the BMC firmware are shown in Figure 4-3 on page 61. Ensure that the **Preserve Configuration** check box is selected and the **Preserve SDR** check box is clear. Click **Start Upgrade**.

Note: You cannot perform other activities that use the BMC interface until the firmware update is complete.

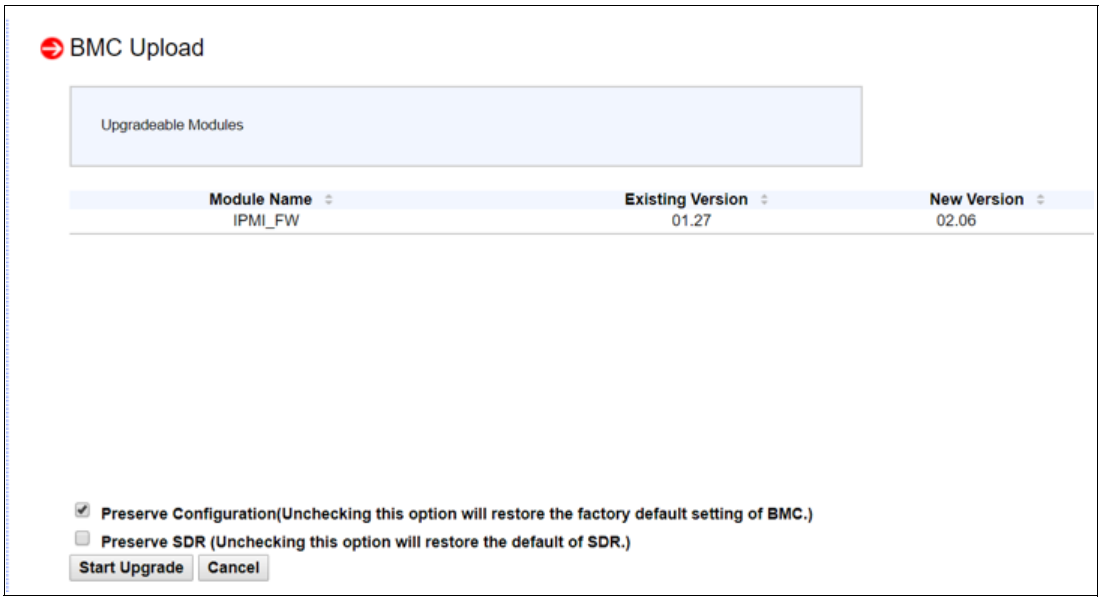


Figure 4-3 BMC Upload: Start Upgrade

6. The upgrade progress of the firmware update is shown in Figure 4-4. After the BMC update is complete, the system restarts.

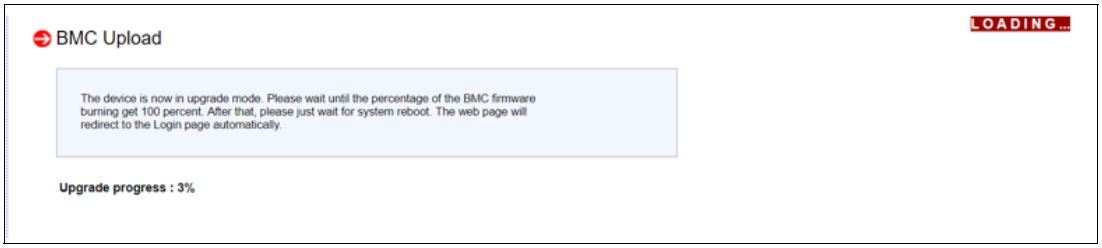


Figure 4-4 BMC Upload window: Update progress

After the restart completes, verify the firmware revision level in the System menu of the BMC dashboard.

To update the PNOR firmware, complete the following steps:

1. Log in to the BMC by entering the user name and password, and then press Enter.
2. From the maintenance list on the dashboard, select **PNOR Update**, as shown in Figure 4-5.

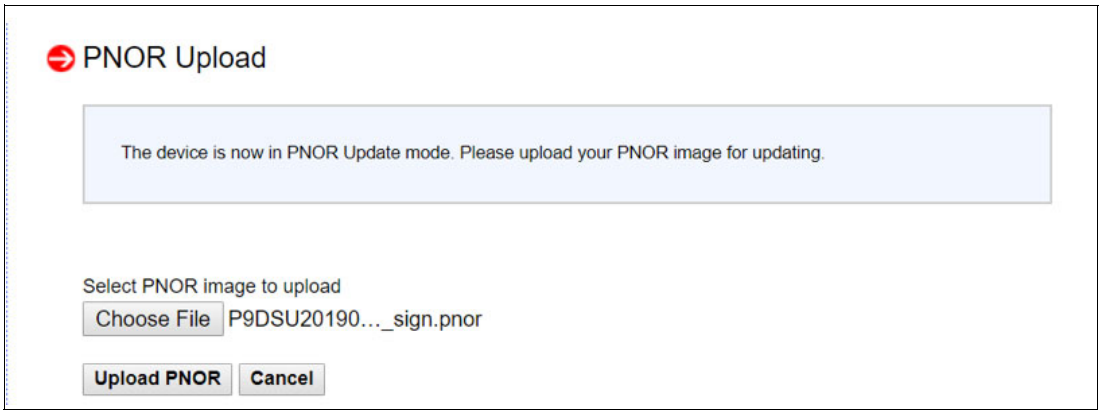


Figure 4-5 PNOR Upload window: Selecting the image to upload

3. In the PNOR Upload window, choose the `.pnor` file from your local system folder and click **Upload PNOR**. Wait for the file to be uploaded, and then click **OK**.
4. The existing and new dates of the PNOR firmware are shown. Click **Start Upgrade**.

> **Note:** You cannot perform other activities by using the BMC interface until the PNOR update is complete.

5. The progress of the PNOR update is shown in Figure 4-6. After the PNOR update completes, the system must be restarted to finish installation of the new PNOR firmware.

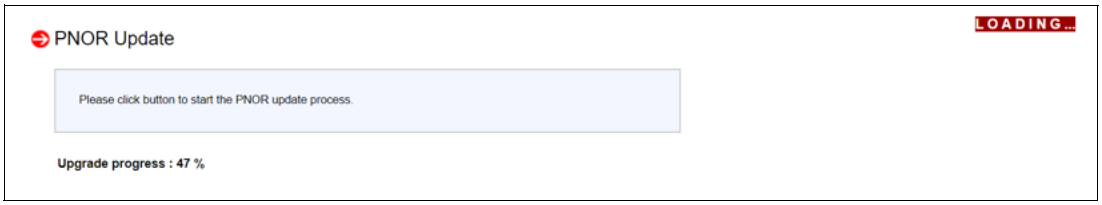


Figure 4-6 PNOR Update window: Upgrade progress

For more information about updating the firmware by using the BMC, see IBM Knowledge Center.

4.1.8 IBM Power RAID Configuration Utility

The following section describes how to configure the internal disks by using the IBM Power RAID Controller Configuration Utility (**iprconfig**). Complete the following steps:

1. In the Petitboot menu, select `Exit to shell`, as shown in Figure 4-7.

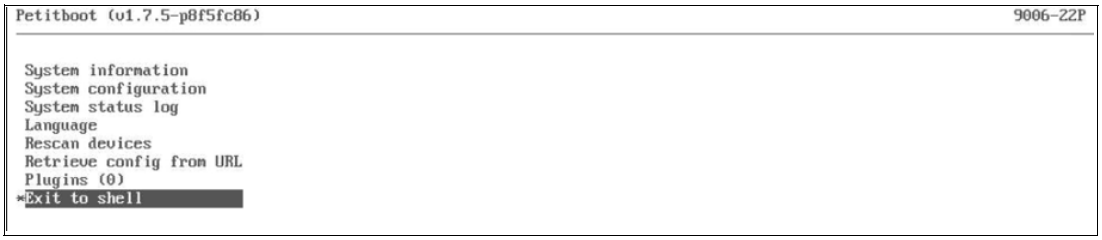


Figure 4-7 Petitboot menu: Exit to shell

2. At the prompt, enter the **iprconfig** command. The following menu opens, as shown in Figure 4-8.

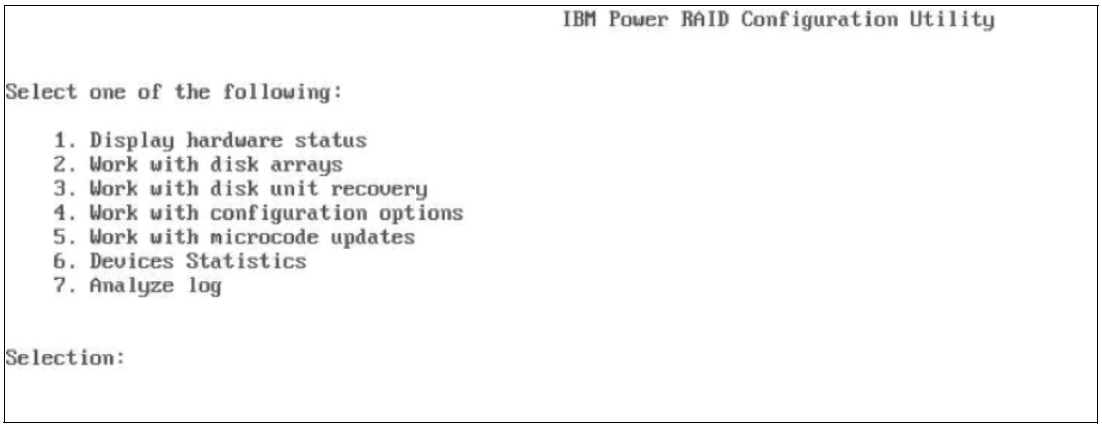


Figure 4-8 IBM Power RAID Configuration Utility menu

3. Enter option 2 to work with disk arrays, as shown in Figure 4-9.

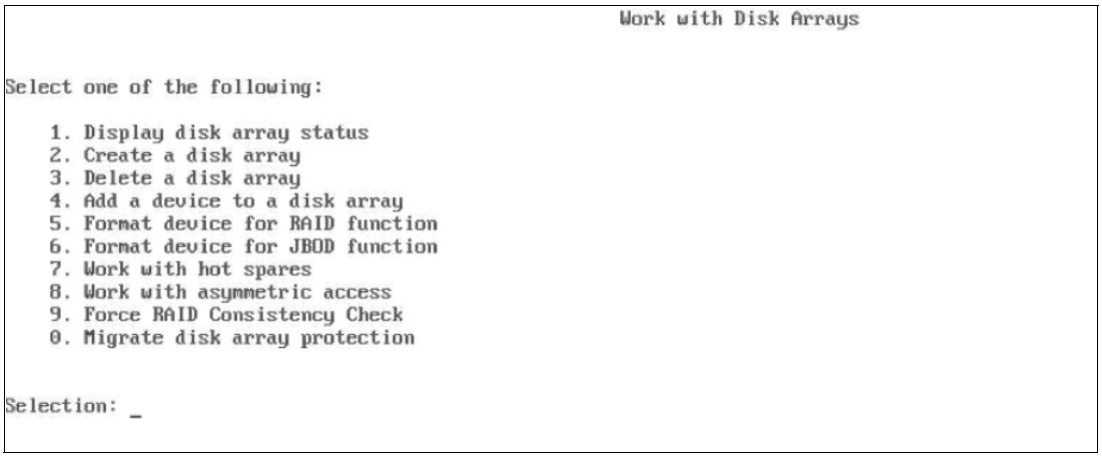


Figure 4-9 Work with Disk Arrays

4.1.9 Adaptec RAID Controller CLI

The following section describes how to configure the internal disks with the Adaptec RAID Controller (**arcconf**). Complete the following steps:

1. In the Petitboot menu, select `Exit to shell`, as shown in Figure 4-10.

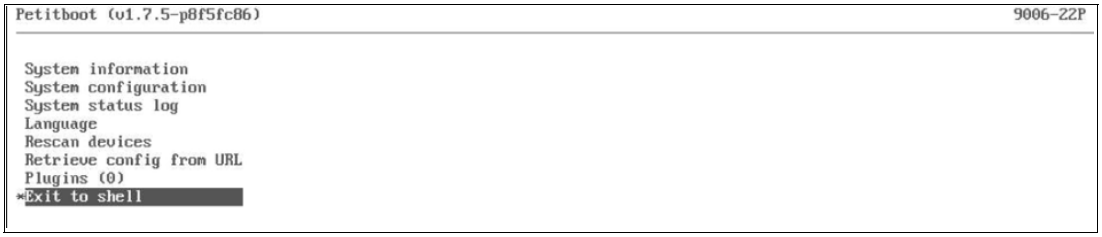


Figure 4-10 Petitboot menu window: Exit to shell

2. At the prompt, run the **arcconf** command and the UCLI pane opens, as shown in Figure 4-11.

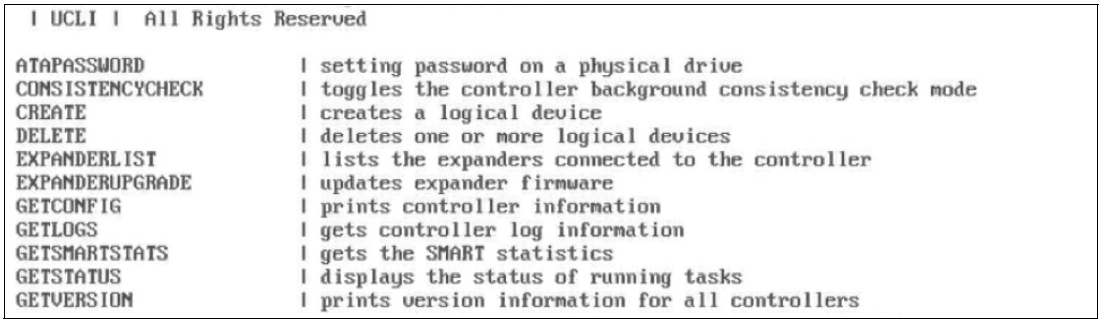


Figure 4-11 UCLI pane: Help window

3. Check the physical disks before you create the arrays by running the following command:

   ```
   # arcconf getconfig 1 PD
   ```

4. To make the output more compact, run the following command:

   ```
   # arcconf getconfig 1 pd|egrep "Device #|State\>|Reported Location|Reported
   Channel| S.M.A.R.T. warnings"
   ```

5. To delete a logical drive, run the following command:

   ```
   arcconf DELETE <Controller#> LOGICALDRIVE <ld#>
   ```

   – For example, to delete logical drive 0, run the following command:

     ```
     # arcconf DELETE 1 LOGICALDRIVE 0
     ```

   – To delete all the drives, run the following command:

     ```
     # arcconf DELETE 1 LOGICALDRIVE ALL
     ```

6. To set up a hardware-based RAID, create a logical drive by running the following command:

   ```
   CREATE <Controller#> LOGICALDRIVE [Options] <Size> <RAID#> <Channel# ID#>
   [Channel# ID#]
   ```

   – To create a RAID 0 with a maximum size; drives on Channel 0; ports 0, 1, 2, and 3; and no confirmation, run the following command:

     ```
     # arcconf CREATE 1 LOGICALDRIVE MAX 0 0 0 0 1 0 2 0 3 noprompt
     ```

   – To create a RAID 5 with a maximum size; drives on Channel 0; ports 0, 1, 2, and 3; and no confirmation, run the following command:

     ```
     # arcconf CREATE 1 LOGICALDRIVE MAX 5 0 0 0 1 0 2 0 3 noprompt
     ```

   – To create a RAID 10 with a maximum size; drives on Channel 0; ports 0, 1, 2, and 3; and no confirmation, run the following command:

     ```
     # arcconf CREATE 1 LOGICALDRIVE MAX 10 0 0 0 1 0 2 0 3 noprompt
     ```

   – To create a RAID 6 with a maximum size; drives on Channel 0; slots 0 - 12; and no confirmation, run the following command:

     ```
     # arcconf CREATE 1 LOGICALDRIVE MAX 6 0 0 0 1 0 2 0 3 0 4 0 5 0 6 0 7 0 8 0
     9 0 10 0 11 0 12 noprompt
     ```

7. To get the logical drive information, run the following command:

   ```
   # arcconf getconfig 1 LD
   ```

4.1.10 Supermicro RAID Controller configuration and firmware update

The section describes how to configure the internal disks with the Supermicro RAID Controller (**`storcli`**). For more information, see IBM Knowledge Center.

This controller type is for use with the PCIe adapter feature codes EKAA and EKEA.

If the **`storcli`** utility command is not installed, download the utility from Broadcom. It can be found by searching the site for the phrase “storcli”. Download the MegaRAID Storcli. To install the driver, extract the file and follow the instructions for the correct OS:

- Red Hat Enterprise Linux:

  a. Go to the `linux-ppc` folder, open it, and then go to the Little Endian folder.

  b. Install the package by running the following command, where *`x.xx-x`* equals the version of the utility:

     ```
     rpm -ivh <StorCLI-x.xx-x.noarch.rpm>
     ```

- Ubuntu:

  a. Go to the Ubuntu folder.

  b. Install the Debian file by running the following command:

     ```
     dpkg -i storcli_x.xx-x._all.deb
     ```

- Preboot and Petitboot

  Note: After a restart, the following steps must be performed again.

  a. From the Petitboot menu, select `Exit to shell`.

  b. Set the IP address if no DHCP server is available in the network by running the following commands:

  ```
  ifconfig <nic-device> <ip-address> netmask <netmask> up
  route add default gw <gw_address> <nic-device>
  ```

  c. Download the `storcli64` binary file into the Petitboot environment by running the following command:

  ```
  wget
  "ftp://ftp.supermicro.com/Firmware/Openpower/P9DSU-C/Storage/Broadcom/AOC-93
  61-8i (AOC-K-9361-8Is-IB001)/007.0606.0000.0000_Unified_StorCLI.zip"
  ```

  d. Extract the following files:

  - `007.0606.0000.0000_Unified_StorCLI.zip`
  - `MR_SAS_Unified_StorCLI_7.6-007.0606.0000.0000-SCGCQ01639776\ \(1\).zip`

  e. Run the following command.

  ```
  cd versionChangeSet/univ_viva_cli_rel/
  ```

  f. Extract the `Unified_storcli_all_os.zip` file by running the following command:

  ```
  cd Unified_storcli_all_os/Linux-PPC/LittleEndian
  ```

  g. Extract the `storcli64.zip` file by running the following command:

  ```
  unzip storcli64.zip
  ```

  h. The `storcli64` file can now be launched by specifying the current directory:

  ```
  ./storcli64 -h
  ```

For more information about managing RAID arrays, see the *12Gb/s MegaRAID SAS Software User Guide*.

Helpful storcli commands

In the following commands, use **`page=[x]`** as the last option to set the page break:

- To show a summary of the drive and controller status, run the following command:

  ```
  ./storcli64 show page=20
  ```

- To show a list of all controllers and drives that need attention, run the following command:

  ```
  ./storcli64 show all page=20
  ```

  Note:

  - **`/c[x]`** or **`/cALL`**: Controller number
  - **`/v[x]`** or **`/vALL`**: VirtualDrive number
  - **`/e[x]`** or **`/eALL`**: Enclosure ID
  - **`/s[x]`** or **`/sALL`**: Slot ID

- To view the adapter settings, run the following command:

  ```
  ./storcli64 /c0 show personality
  ```

- To change the adapter settings, run the following command:

  ```
  ./storcli64 /c0 set personality=JBOD|RAID
  ```

Updating firmware

Here are useful commands to use when updating the firmware:

- To check the firmware level of the MegaRAID adapter, run the following command:

  ```
  ./storcli64 /c0 show all|egrep  "Bios Version|Firmware Package|Firmware
  Version"
  ```

- To update the firmware level of the MegaRAID adapter, run the following command:

  ```
  ./storcli64 /c[x] download file=mrxxxxfw.rom [ResetNow]
  ```

  Note: The **ResetNow** option is necessary when the old firmware is still active after flushing the adapter.

  Figure 4-12 shows an example without the **ResetNow** option.

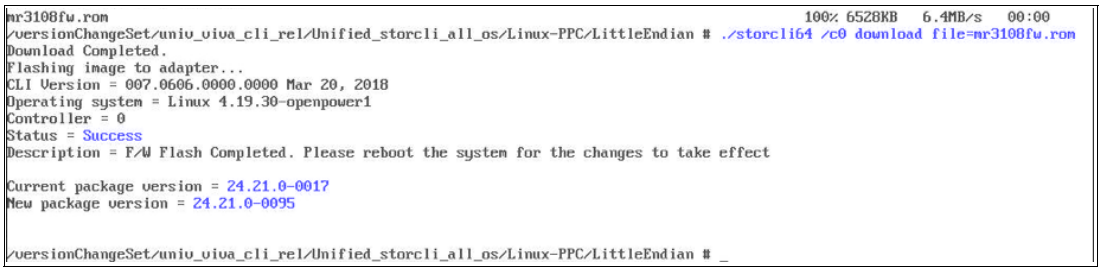


Figure 4-12 Command to flash the controller firmware without the ResetNow option

Figure 4-13 shows the command to flash the firmware with the **ResetNow** option.

The following command is used to flash the controller firmware.

storcli /cx download file=*filepath* [fwtype=<*value*>] [nosigchk] [noverchk] [resetnow]

This command flashes the firmware with the ROM file to the specified adapter from the given file location (*filepath* is the absolute file path).

You can use the following options in the table to flash the firmware:

Table 9 Flashing Controller Firmware Input Options

Option	Value Range	Description
nosigchk	—	The application flashes the firmware even if the check word on the file does not match the required check word for the controller. **NOTE** You can damage the controller if a corrupted image is flashed using this option.

Table 9 Flashing Controller Firmware Input Options (Continued)

Option	Value Range	Description
noverchk	—	The application flashes the controller firmware without checking the version of the firmware image.
fwtype	0: Application	The firmware type to be downloaded. The application downloads the firmware for the controller.
resetnow	—	Invokes online firmware update on the controller; you do not need to reboot the controller to make the update effective.

Figure 4-13 Command to flash the controller firmware with the ResetNow option

- To show information about the drives, run the following command:

  `./storcli64 /c[x](/eALL)/sALL show`

  For example (Figure 4-14):

  `./storcli64 /c0/eALL/sALL show`

```
/versionChangeSet/univ_viva_cli_rel/Unified_storcli_all_os/Linux-PPC/LittleEndian # ./storcli64 /c0/eALL/sALL show all page=20
CLI Version = 007.0606.0000.0000 Mar 20, 2018
Operating system = Linux 4.19.30-openpower1
Controller = 0
Status = Success
Description = Show Drive Information Succeeded.


Drive /c0/e8/s1 :
================

-------------------------------------------------------------------------------
EID:Slt DID State DG     Size Intf Med SED PI SeSz Model                Sp Type
-------------------------------------------------------------------------------
8:1      18 Onln   0 3.637 TB SATA HDD N   N  512B ST4000NM0115-1YZ107 U  -
-------------------------------------------------------------------------------

EID-Enclosure Device ID|Slt-Slot No.|DID-Device ID|DG-DriveGroup
DHS-Dedicated Hot Spare|UGood-Unconfigured Good|GHS-Global Hotspare
UBad-Unconfigured Bad|Onln-Online|Offln-Offline|Intf-Interface
Med-Media Type|SED-Self Encryptive Drive|PI-Protection Info
Press any key to continue
```

Figure 4-14 Showing information about the drives

Figure 4-15 shows the detailed drive information.

```
Drive /c0/e8/s1 - Detailed Information :
=======================================

Drive /c0/e8/s1 State :
=======================
Shield Counter = 0
Media Error Count = 0
Other Error Count = 0
Drive Temperature =  32C (89.60 F)
Predictive Failure Count = 0
S.M.A.R.T alert flagged by drive = No


Drive /c0/e8/s1 Device attributes :
===================================
Press any key to continue
SN = ZC19RD4K
Manufacturer Id = ATA
Model Number = ST4000NM0115-1YZ107
NAND Vendor = NA
WWN = 5000c500b42100c1
Firmware Revision = SN04
Raw size = 3.638 TB [0x1d1c0beb0 Sectors]
Coerced size = 3.637 TB [0x1d1b00000 Sectors]
Non Coerced size = 3.637 TB [0x1d1b0beb0 Sectors]
Device Speed = 6.0Gb/s
Link Speed = 12.0Gb/s
NCQ setting = Enabled
Write Cache = N/A
Logical Sector Size = 512B
Physical Sector Size = 4 KB
Connector Name = Port 0 - 3 & Port 4 - 7


Drive /c0/e8/s1 Policies/Settings :
===================================
Press any key to continue
```

Figure 4-15 Showing detailed drive information

- To find a physical disk by turning on the identify LED, run the following command:

  `./storcli64 /c[x](/e[x])/s[x] start locate`

- To turn off the identify LED, run the following command:

  `./storcli64 /c[x](/e[x])/s[x] stop locate`

- To show the predictive failure count (Figure 4-16), run the following command:

  `./storcli64 /c0/eALL/sALL show all | grep -e 'State :' -e "Predictive Failure Count" -e "HDD"`

```
/versionChangeSet/univ_viva_cli_rel/Unified_storcli_all_os/Linux-PPC/LittleEndian # ./storcli64 /c0/eALL/sALL show all | grep -e
 'State :' -e "Predictive Failure Cout" -e "HDD"
8:1      18 Onln   0 3.637 TB SATA HDD N   N  512B ST4000NM0115-1YZ107 U  -
Drive /c0/e8/s1 State :
8:2      20 Onln   0 3.637 TB SATA HDD N   N  512B ST4000NM0115-1YZ107 U  -
Drive /c0/e8/s2 State :
8:9       9 Onln   0 3.637 TB SATA HDD N   N  512B ST4000NM0115-1YZ107 U  -
Drive /c0/e8/s9 State :
8:10     11 Onln   0 3.637 TB SATA HDD N   N  512B ST4000NM0115-1YZ107 U  -
Drive /c0/e8/s10 State :
/versionChangeSet/univ_viva_cli_rel/Unified_storcli_all_os/Linux-PPC/LittleEndian #
```

Figure 4-16 Showing the predictive failure count

- To show information about virtual drives (Figure 4-17), run the following command:

  `./storcli64 /c0/vALL show`

```
/versionChangeSet/univ_viva_cli_rel/Unified_storcli_all_os/Linux-PPC/LittleEndian # ./storcli64 /c0/vALL show page=40
CLI Version = 007.0606.0000.0000 Mar 20, 2018
Operating system = Linux 4.19.30-openpower1
Controller = 0
Status = Success
Description = None


Virtual Drives :
==============

-------------------------------------------------------------
DG/VD TYPE  State Access Consist Cache Cac sCC     Size Name
-------------------------------------------------------------
0/0   RAID0 Optl  RW     Yes     RWBD  -   ON  3.637 TB
1/1   RAID6 Optl  RW     No      RWBD  -   ON  3.637 TB
-------------------------------------------------------------
```

Figure 4-17 Showing virtual drive information

  - The following commands create two virtual drives:
    - **`./storcli64 /c0 add vd type=r0 drives=8:1-2`**
    - **`./storcli64 /c0 add vd type=r6 drives=8:2,8-9`**
  - To create a virtual drive, use the following command:

    `./storcli64 /c[x] add vd type=[RAID0(r0)|RAID1(r1)|...]`
    `drives=[EnclosureID:SlotID|:SlotID-SlotID|:SlotID,SlotID]`
  - To create a RAID 0 array (Figure 4-18), run the following command:

    `./storcli64 /c0 add vd type=r0 drives=8:1-2`

```
/versionChangeSet/univ_viva_cli_rel/Unified_storcli_all_os/Linux-PPC/LittleEndian # ./storcli64 /c0/vall show all page=40
CLI Version = 007.0606.0000.0000 Mar 20, 2018
Operating system = Linux 4.19.30-openpower1
Controller = 0
Status = Success
Description = None


/c0/v0 :
======

-------------------------------------------------------------
DG/VD TYPE  State Access Consist Cache Cac sCC     Size Name
-------------------------------------------------------------
0/0   RAID0 Optl  RW     Yes     RWBD  -   ON  7.276 TB
-------------------------------------------------------------

Cac=CacheCade|Rec=Recovery|OfLn=OffLine|Pdgd=Partially Degraded|Dgrd=Degraded
Optl=Optimal|RO=Read Only|RW=Read Write|HD=Hidden|TRANS=TransportReady|B=Blocked|
Consist=Consistent|R=Read Ahead Always|NR=No Read Ahead|WB=WriteBack|
AWB=Always WriteBack|WT=WriteThrough|C=Cached IO|D=Direct IO|sCC=Scheduled
Check Consistency


PDs for VD 0 :
============

-----------------------------------------------------------------------------
EID:Slt DID State DG     Size Intf Med SED PI SeSz Model                Sp Type
-----------------------------------------------------------------------------
8:1      18 Onln   0 3.637 TB SATA HDD N   N  512B ST4000NM0115-1YZ107 U  -
8:2      20 Onln   0 3.637 TB SATA HDD N   N  512B ST4000NM0115-1YZ107 U  -
-----------------------------------------------------------------------------
```

Figure 4-18 Showing the RAID 0 configuration

– To create a RAID 1 array (Figure 4-19), run the following command:

```
./storcli64 /c0 add vd type=r1 drives=8:1-2
```

```
/versionChangeSet/univ_viva_cli_rel/Unified_storcli_all_os/Linux-PPC/LittleEndian # ./storcli64 /c0/vall show all page=40
CLI Version = 007.0606.0000.0000 Mar 20, 2018
Operating system = Linux 4.19.30-openpower1
Controller = 0
Status = Success
Description = None


/c0/v0 :
======

-------------------------------------------------------------
DG/VD TYPE  State Access Consist Cache Cac sCC     Size Name
-------------------------------------------------------------
0/0   RAID1 Optl  RW     No      RWBD  -   ON  3.637 TB
-------------------------------------------------------------

Cac=CacheCade|Rec=Recovery|OfLn=OffLine|Pdgd=Partially Degraded|Dgrd=Degraded
Optl=Optimal|RO=Read Only|RW=Read Write|HD=Hidden|TRANS=TransportReady|B=Blocked|
Consist=Consistent|R=Read Ahead Always|NR=No Read Ahead|WB=WriteBack|
AWB=Always WriteBack|WT=WriteThrough|C=Cached IO|D=Direct IO|sCC=Scheduled
Check Consistency


PDs for VD 0 :
============

-------------------------------------------------------------------------
EID:Slt DID State DG     Size Intf Med SED PI SeSz Model            Sp Type
-------------------------------------------------------------------------
8:1      18 Onln   0 3.637 TB SATA HDD N   N  512B ST4000NM0115-1YZ107 U  -
8:2      20 Onln   0 3.637 TB SATA HDD N   N  512B ST4000NM0115-1YZ107 U  -
-------------------------------------------------------------------------
```

Figure 4-19 Showing the RAID 1 configuration

– To create a RAID 5 array (Figure 4-20), run the following command:

```
./storcli64 /c0 add vd type=r5 drives=8:1-2,8-9
```

```
/versionChangeSet/univ_viva_cli_rel/Unified_storcli_all_os/Linux-PPC/LittleEndian # ./storcli64 /c0/vALL show page=40
CLI Version = 007.0606.0000.0000 Mar 20, 2018
Operating system = Linux 4.19.30-openpower1
Controller = 0
Status = Success
Description = None


Virtual Drives :
==============

-------------------------------------------------------------
DG/VD TYPE  State Access Consist Cache Cac sCC      Size Name
-------------------------------------------------------------
0/0   RAID5 Optl  RW     No      RWBD  -   ON  10.914 TB
-------------------------------------------------------------
```

Figure 4-20 Showing the RAID 5 configuration

– To create a RAID 6 array (Figure 4-21), run the following command:

```
./storcli64 /c0 add vd type=r6 drives=8:1,2,8,9
```

```
/versionChangeSet/univ_viva_cli_rel/Unified_storcli_all_os/Linux-PPC/LittleEndian # ./storcli64 /c0/vall show all page=40
CLI Version = 007.0606.0000.0000 Mar 20, 2018
Operating system = Linux 4.19.30-openpower1
Controller = 0
Status = Success
Description = None

/c0/v0 :
======

-----------------------------------------------------------------
DG/VD TYPE  State Access Consist Cache Cac sCC     Size Name
-----------------------------------------------------------------
0/0   RAID6 Optl  RW     No      RWBD  -   ON  7.276 TB
-----------------------------------------------------------------

Cac=CacheCade|Rec=Recovery|OfLn=OffLine|Pdgd=Partially Degraded|Dgrd=Degraded
Optl=Optimal|RO=Read Only|RW=Read Write|HD=Hidden|TRANS=TransportReady|B=Blocked|
Consist=Consistent|R=Read Ahead Always|NR=No Read Ahead|WB=WriteBack|
AWB=Always WriteBack|WT=WriteThrough|C=Cached IO|D=Direct IO|sCC=Scheduled
Check Consistency

PDs for VD 0 :
============

-------------------------------------------------------------------------------
EID:Slt DID State DG     Size Intf Med SED PI SeSz Model                Sp Type
-------------------------------------------------------------------------------
8:1      18 Onln   0 3.637 TB SATA HDD N   N  512B ST4000NM0115-1YZ107 U  -
8:2      20 Onln   0 3.637 TB SATA HDD N   N  512B ST4000NM0115-1YZ107 U  -
8:9       9 Onln   0 3.637 TB SATA HDD N   N  512B ST4000NM0115-1YZ107 U  -
8:10     11 Onln   0 3.637 TB SATA HDD N   N  512B ST4000NM0115-1YZ107 U  -
-------------------------------------------------------------------------------
```

Figure 4-21 Showing the RAID 6 configuration

If a more complex RAID (for example, RAID10) is required, it is necessary to specify the number of disks per RAID by running the following command:

```
./storcli64 /c[x] add vd type=r[x] drives=8:0-3 PDperArray=2
```

- To initialize a virtual drive, the command syntax is as follows:

  ```
  ./storcli64 /c[x]/v[x] start init (force)
  ```

  For example (Figure 4-22):

  ```
  ./storcli64 /c0/v0 start init
  ```

```
/versionChangeSet/univ_viva_cli_rel/Unified_storcli_all_os/Linux-PPC/LittleEndian # ./storcli64 /c0/v0 start init
CLI Version = 007.0606.0000.0000 Mar 20, 2018
Operating system = Linux 4.19.30-openpower1
Controller = 0
Status = Success
Description = Start INIT Operation Success
```

Figure 4-22 Initializing the drive

The progress can be monitored by running the following command (Figure 4-23):

```
./storcli64 /c[x]/v[x] show init or ./storcli64 /c[x]/vALL show init
```

```
/versionChangeSet/univ_viva_cli_rel/Unified_storcli_all_os/Linux-PPC/LittleEndian # ./storcli64 /c0/v0 show init page=40
CLI Version = 007.0606.0000.0000 Mar 20, 2018
Operating system = Linux 4.19.30-openpower1
Controller = 0
Status = Success
Description = None

VD Operation Status :
==================

-------------------------------------------------
VD Operation Progress% Status          Estimated Time Left
-------------------------------------------------
 0 INIT      -         Not in progress -
-------------------------------------------------
```

Figure 4-23 Monitoring the initialization progress

- To delete virtual drives, use the following command syntax:

  `./storcli64 /c[x]/v[x] del (force)`

  For example:

  `./storcli64 /c0/vALL del`

- To prepare a drive for removal, use the following command syntax:

  `./storcli64 /c[x](/e[x])/s[x] spindown`

4.2 Loading and initializing Linux on bare metal Power Systems hardware

This section describes the process to load and initialize the basic Linux OS on local disks.

4.2.1 Attaching a bootable DVD by using the USB device and configuring Petitboot

After the system powers on, the Petitboot bootloader scans local boot devices and network interfaces to find boot options that are available to the system. To use a USB device, complete these steps:

1. Insert your bootable USB device into the front USB port. Petitboot displays the following option:

   ```
   Note: Select Rescan devices if the USB device does not appear. If your device
   is not detected, you have to try a different device type.
   ```

2. Record the UUID of the USB device. For example, the UUID of the USB device in the following example is 2015-10-30-11-05-03-00.

   ```
   [USB:sdb1 / 2015-10-30-11-05-03-00]
             Rescue a Red Hat Enterprise Linux system (64-bit kernel)
             Test this media & install Red Hat Enterprise Linux 7.x (64-bit kernel)
         *   Install Red Hat Enterprise Linux 7.x (64-bit kernel)
   ```

3. Select `Install Red Hat Enterprise Linux 7.x (64-bit kernel)` and select `e` (Edit) to open the Petitboot Option Editor window.

4. Move the cursor to the boot arguments section, and add the following information:

   `inst.text inst.stage2=hd:UUID=your_UUID`

   The `your_UUID` is the UUID that you recorded.

   ```
   Petitboot Option Editor

   qqqqqqqqqqqqqqqqqqqqqqqqqqqqqqqqqqqqqqqqqqqqqqqqqqqqqqqqqqqqqqqqqqqqqqqqqqqqq
               Device:   ( ) sda2 [f8437496-78b8-4b11-9847-bb2d8b9f7cbd]
                             (*) sdb1 [2015-10-30-11-05-03-00]
                             (  ) Specify paths/URLs manually

               Kernel: /ppc/ppc64/vmlinuz
               Initrd: /ppc/ppc64/initrd.img
               Device tree:
               Boot arguments: ro inst.text inst.stage2=hd:UUID=2015-10-30-11-05-03-00
                  [    OK    ]  [   Help   ]  [  Cancel  ]
   ```

5. Select `OK` to save your options and return to the main menu.
6. Verify that `Install Red Hat Enterprise Linux 7.x (64-bit kernel)` is selected, and then press Enter to begin your installation.

4.2.2 Attaching an ISO image by using virtual media (BMC Web Front End)

BMC Advanced Systems Management is a remote management controller that you use to access system information, status, and other processes for your server. You can use the BMC Advanced Systems Management to set up your installation and provide the CD image as virtual media to the Power Systems server. However, the actual installation requires a serial over LAN (SOL) connection through IPMI.

To access the BMC Advanced Systems Management, open a web browser and go to `http://`*`ip_address`*, where `ip_address` is the IP address for the BMC. Log in by using these default values:

- Default user name: ADMIN
- Default password: ADMIN or admin

There are two ways to connect an ISO image:

- Using the Java console
- Using a Windows share

To fully use BMC Advanced Systems Management, the IP address of the BMC firmware must be added to the Exceptions list in the Java Control Panel of your notebook or PC.

On a Windows system, select **Control Panel** → **Control Panel for Java**. After accessing the Control Panel for Java, select the **Security** tab, add the IP address of the BMC firmware to the Exceptions list by clicking **Edit Site List**, and then click **Add**. Enter the IP address and click **OK**.

On a Linux system, select **Control Central** and then select the Java web browser plug-in.

To create a virtual CD/DVD, complete these steps:

1. Log in to the BMC Advanced Systems Management interface from a PC or notebook by using the default user name and password.
2. Select **Remote Control** → **Console Redirection**.
3. Select **Java Console**. As the console opens, you might need to direct your browser to open the `jviewer.jnlp` file by selecting **Open with Java Web Start** and clicking **OK**. Accept the warning and click **Run**.

4. In the Console Redirection window (Figure 4-24), select **Media** → **Virtual Media wizard** from the menu.

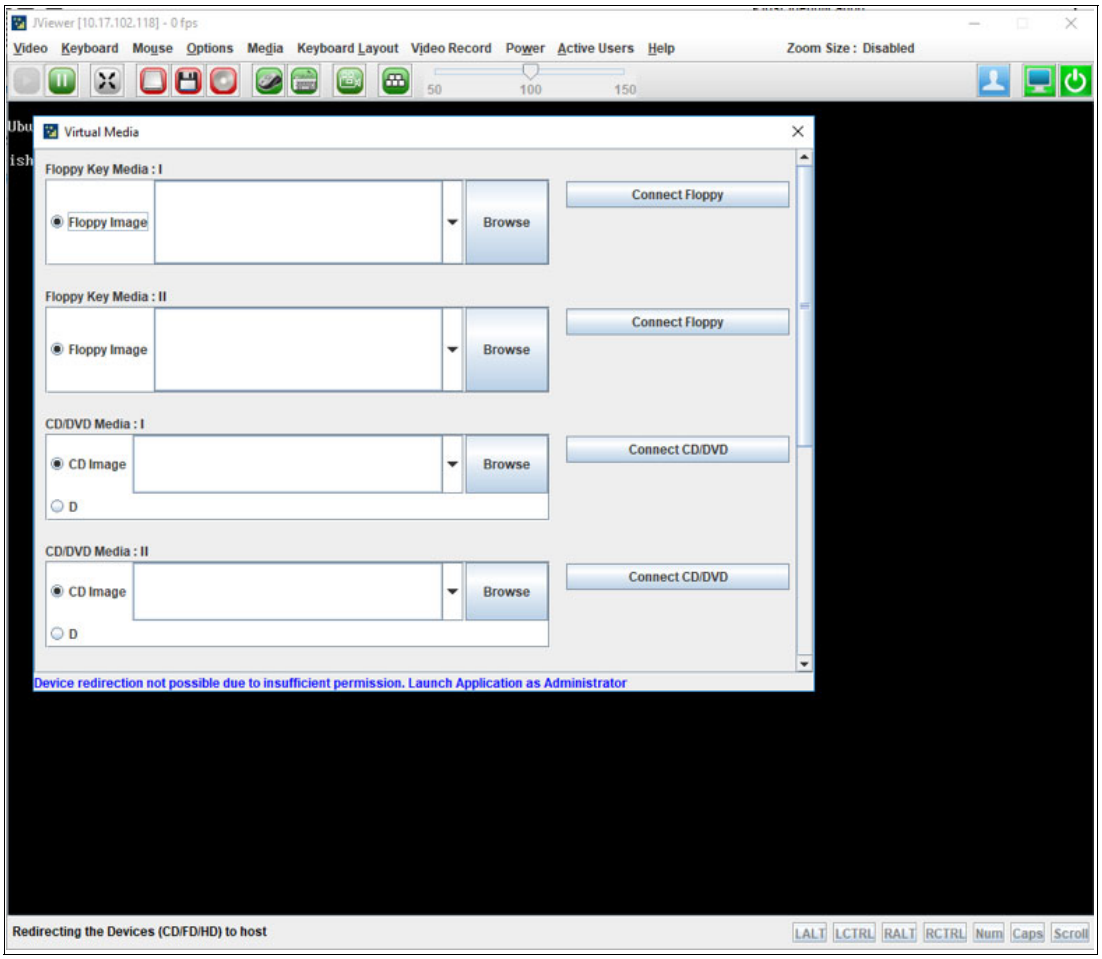


Figure 4-24 Console Redirection window

5. In the Virtual Media wizard, select CD/DVD Media:1. Select **CD Image** and the path to the Linux distribution ISO file. For example, `/tmp/RHEL-7.2-20151030.0-Server-ppc64el-dvd1.iso`.

   Click **Connect CD/DVD**. If the connection is successful, the message “Device redirected in Read Only Mode” appears.

6. Verify that CD/DVD is shown as an option in Petitboot as sr0:

```
CD/DVD:  sr0
  Install
  Repair
```

> **Note:** Select **Rescan devices** if CD/DVD does not appear.

7. Select **Install**. After selecting **Install**, your remote console might become inactive. Open or reactivate your IPMI console to complete the installation.

> **Note:** Be patient. It can sometimes take a couple minutes for the installation to begin.

Attaching an ISO image over a Windows share

With this option, you can share a CD-ROM image over a Windows share with a maximum size of 4.7 GB. This image is emulated to the host as a USB device.

Complete the following steps:

1. Log in to the BMC Advanced Systems Management interface from a PC or notebook by using the default user name and password.
2. Select **Virtual Media** → **CD-ROM Image**.
3. Complete the **Share host**, **Path to image**, **User (optional)**, and **Password (optional)** fields.

> **Note:** Use the IP address if there are problems with name resolution in the network.

4. Click **Save**.
5. Click **Mount**.
6. Click **Refresh Status**, as shown in Figure 4-25.

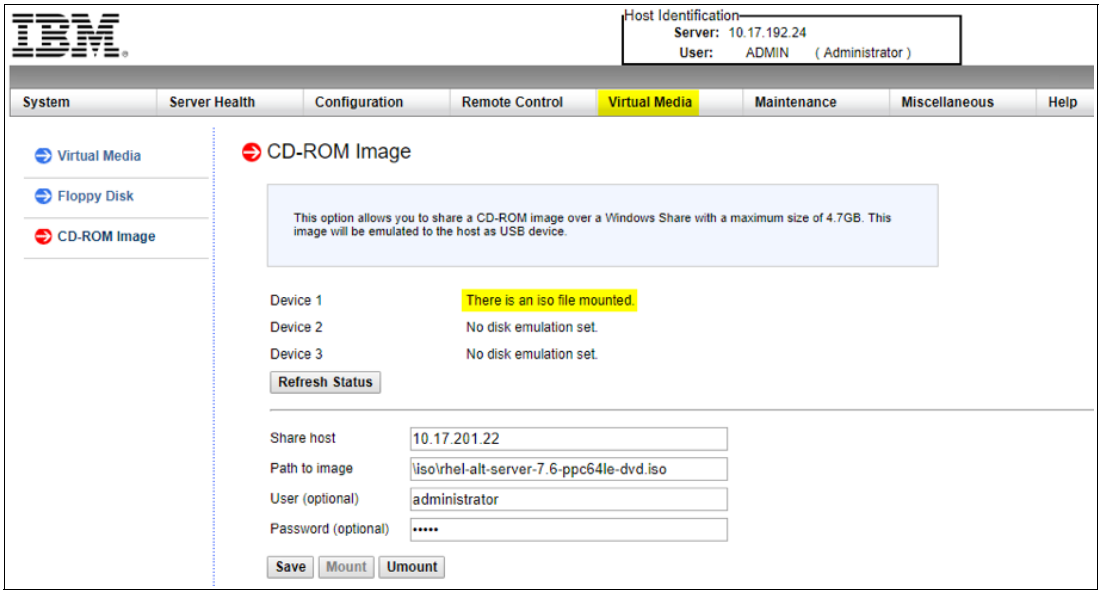


Figure 4-25 BMC Advanced Systems Management interface window

Figure 4-26 shows part of the window that is visible when booting from CD/DVD if the attached ISO image works correctly.

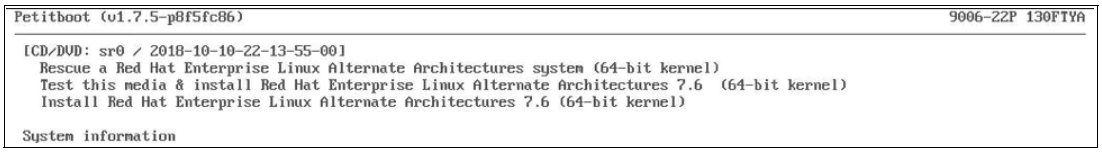


Figure 4-26 Petitboot pane: Booting from CD/DVD

4.3 Installing RHEL OS as a KVM host on bare metal Power Systems hardware

This section contains the RHEL installer options and requirements.

4.3.1 Red Hat Enterprise Linux installer options

For more information about installing RHEL, see *Installation Guide Red Hat Virtualization 4.2*.

Hardware requirements

The minimum and recommended hardware requirements that are shown in Table 4-1 are based on a typical small- to medium-sized installation. The exact requirements vary among deployments based on sizing and load.

Hardware certification for RHV is covered by the hardware certification for RHEL. For more information, see Does Red Hat Virtualization also have hardware certification?

To confirm whether specific hardware items are certified for use with RHEL, see Red Hat Certified Products & Services.

Table 4-1 Red Hat Virtualization Manager hardware requirements

Resource	Recommended
CPU	A quad core CPU or multiple dual-core CPUs.
Memory	16 GB of system RAM.
Hard disk	50 GB of locally accessible, writable disk space. You can use the RHV Manager (RHV-M) History Database Size Calculator to calculate the appropriate disk space for the Manager history database size.
Network interface	One network interface card (NIC) with bandwidth of at least 1 Gbps.

Storage requirements

Hosts require local storage to store configuration, logs, and kernel dumps, and for use as swap space. The minimum storage requirements of RHV hosts are documented in this section. The storage requirements for RHEL hosts vary based on the amount of disk space that is used by their existing configuration, but are expected to be greater than the requirements of the RHV host.

The minimum storage requirements for host installation are as follows. However, Red Hat recommends using the default allocations, which use more storage space.

- / (root): 6 GB.
- `/home`: 1 GB.
- `/tmp`: 1 GB.
- `/boot`: 1 GB.
- `/var`: 15 GB.
- `/var/log`: 8 GB.
- `/var/log/audit`: 2 GB.

- Swap size: 1 GB. (For the recommended swap size, see What is the recommended swap size for Red Hat platforms?)
- Anaconda reserves 20% of the thin pool size within the volume group for future metadata expansion to prevent a ready for use configuration from running out of space under normal usage conditions. Overprovisioning of thin pools during installation is not supported.
- Minimum total: 45 GB.

If you are also installing the RHV-M Appliance for self-hosted engine installation, `/var/tmp` must be at least 5 GB.

4.3.2 Operating system requirements

The RHV-M must be installed on a base installation of RHEL 7 that is updated to the latest minor release. To update RHEL, complete the following steps.

Important: Do not install any extra packages after the base installation because they can cause dependency issues when attempting to install the packages that are required by the Virtualization Manager.

Do not enable more repositories other than the ones that are required for the Virtualization Manager installation.

1. Set the disk layout by clicking **INSTALLATION DESTINATION**, as shown in Figure 4-27.

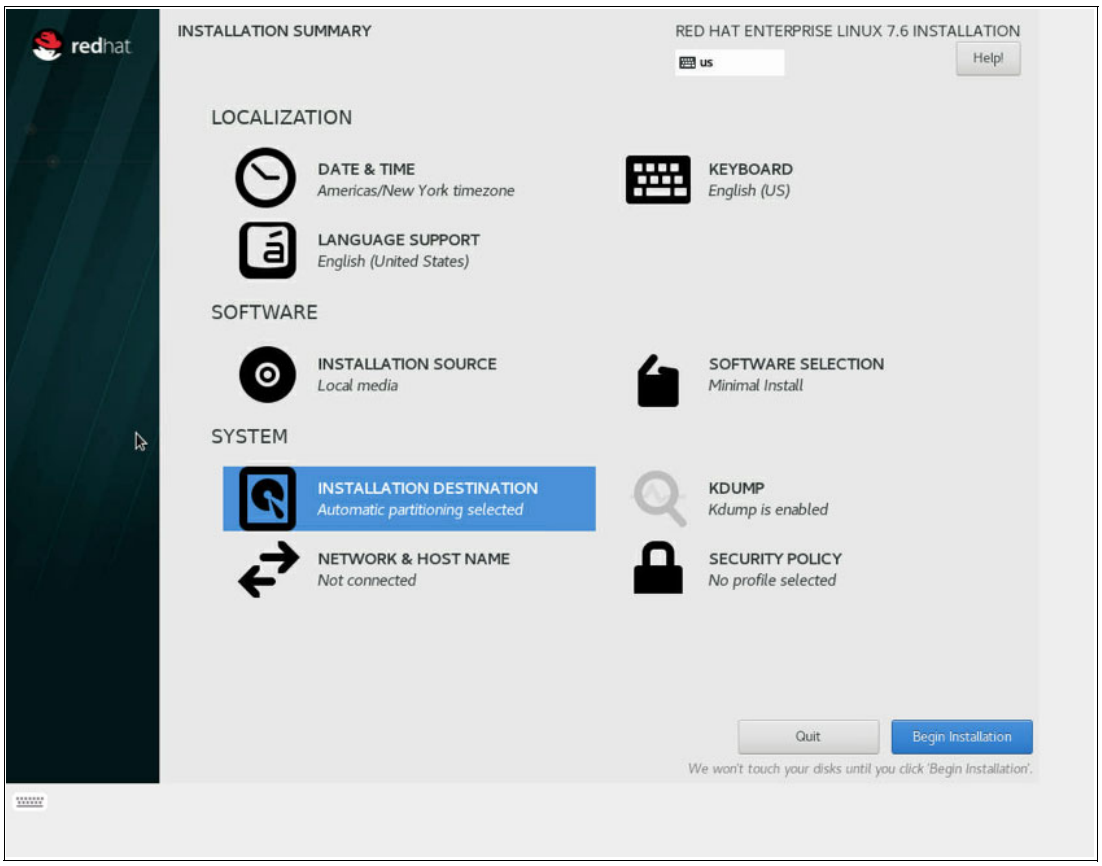


Figure 4-27 Red Hat Enterprise Linux 7.6 Installation window: INSTALLATION DESTINATION

2. Click **Done**, as shown in Figure 4-28.

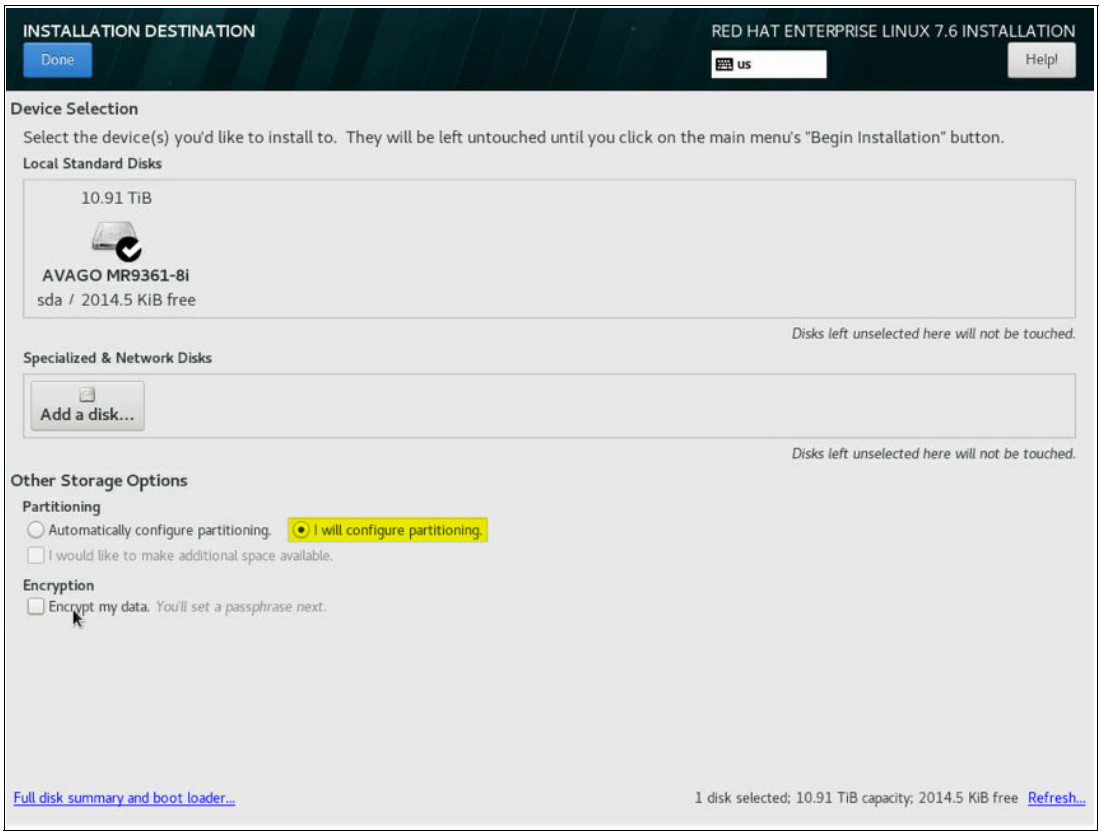


Figure 4-28 Red Hat Enterprise Linux 7.6 Installation window: Device Selection

Figure 4-29 shows the selected disk.

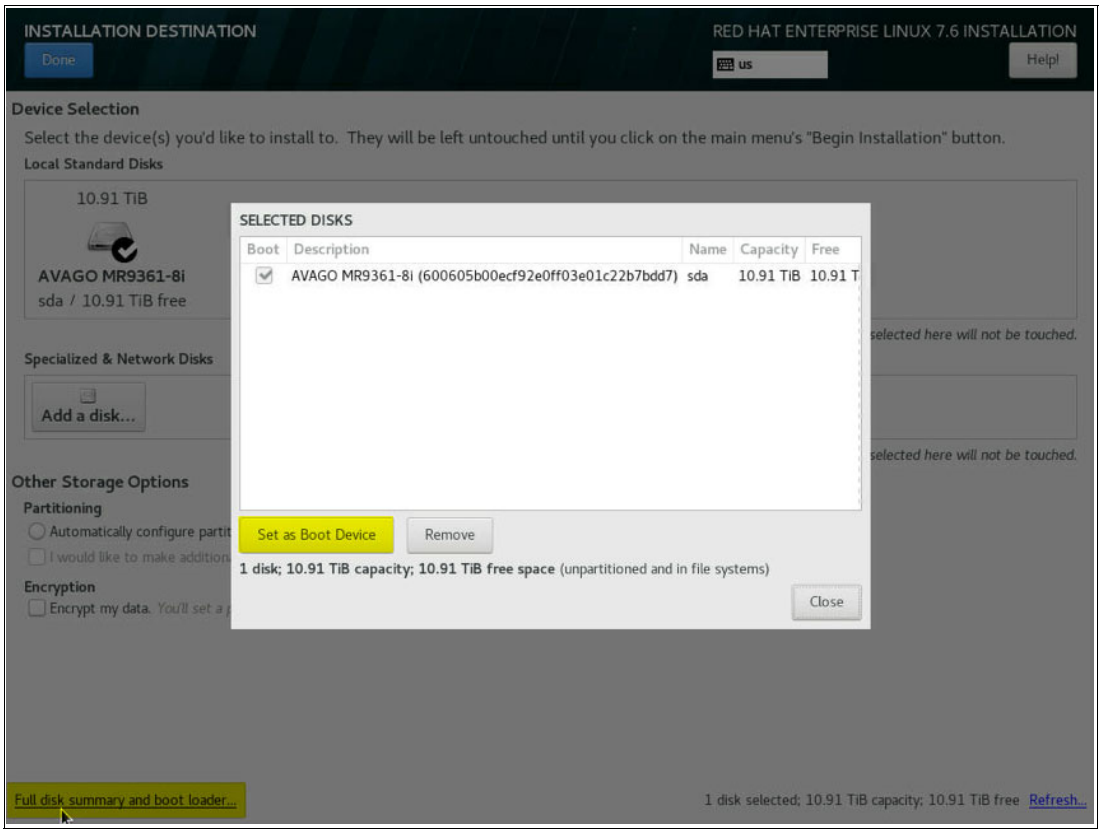


Figure 4-29 Red Hat Enterprise Linux 7.6 Installation window: Selected disk

Figure 4-30 shows how to create manually the mounting point for the installation.

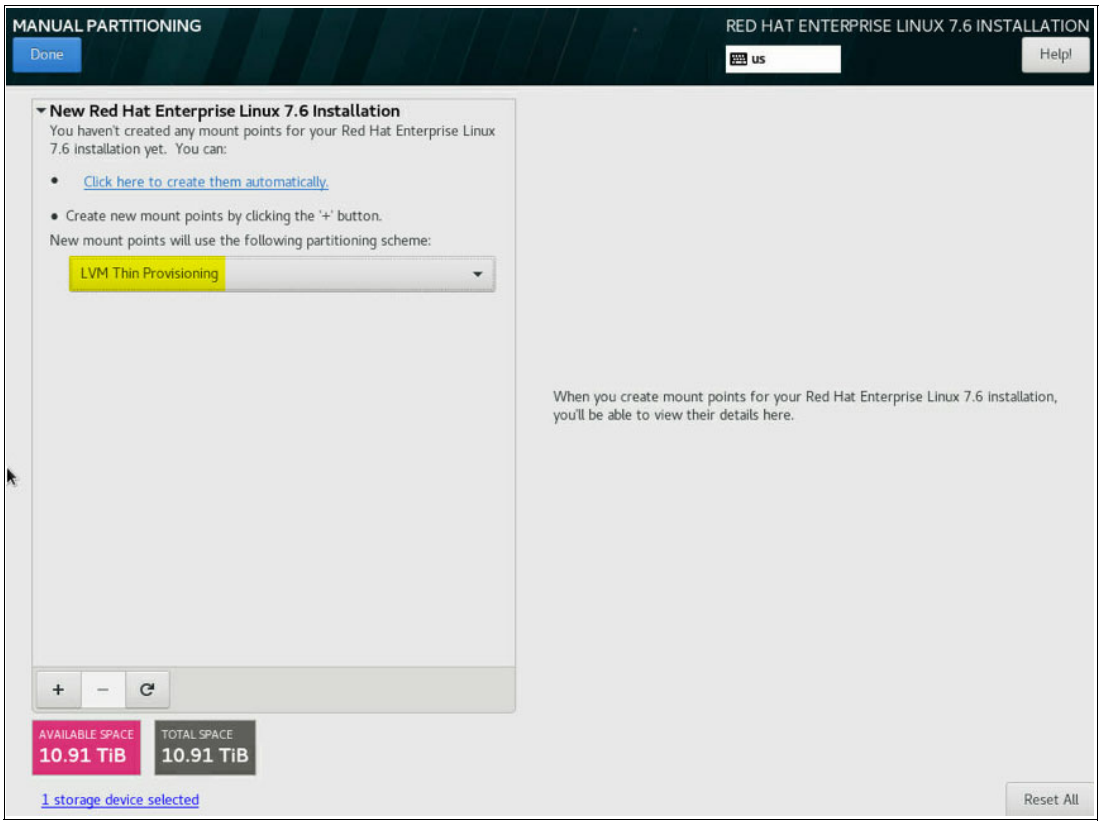


Figure 4-30 Red Hat Enterprise Linux 7.6 Installation window: Manual partitioning

Figure 4-31 shows the window for customizing the disk layout for the installation.

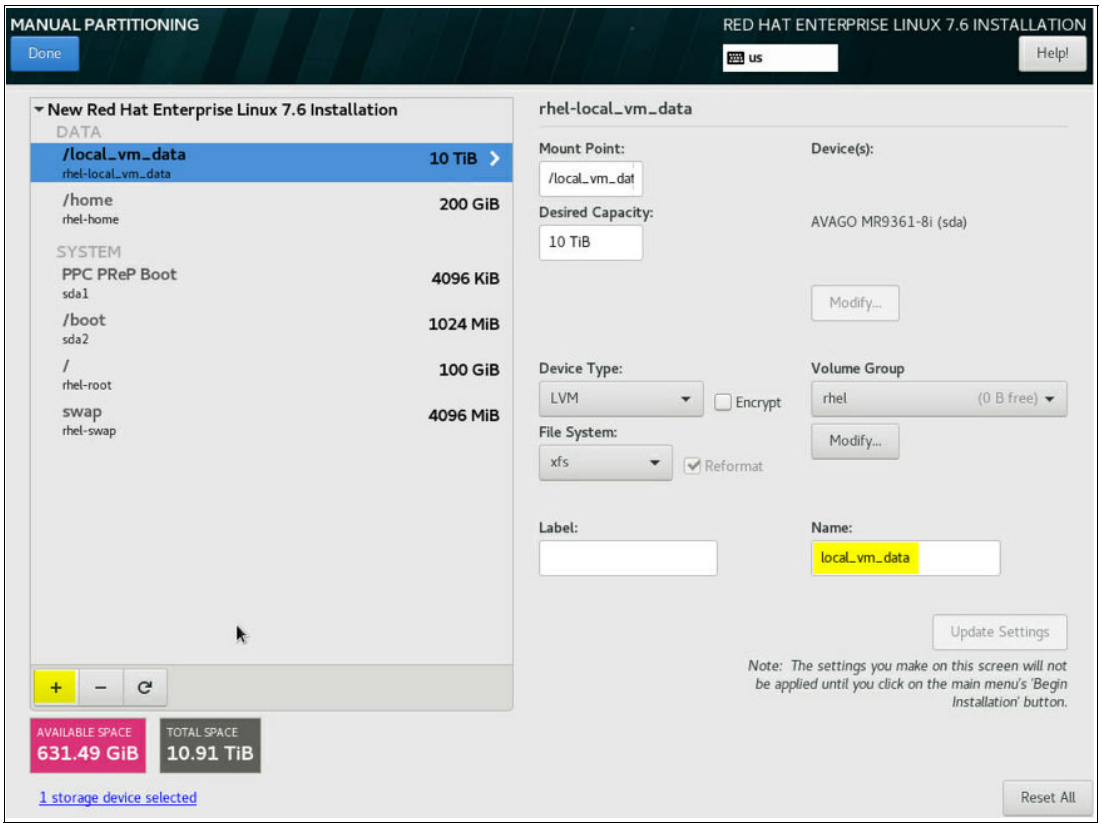


Figure 4-31 Red Hat Enterprise Linux 7.6 Installation window: Customizing the disk layout

3. Click **Done** (Figure 4-31 on page 82), and accept the changes that are shown in the INSTALLATION SUMMARY, as shown in Figure 4-32.

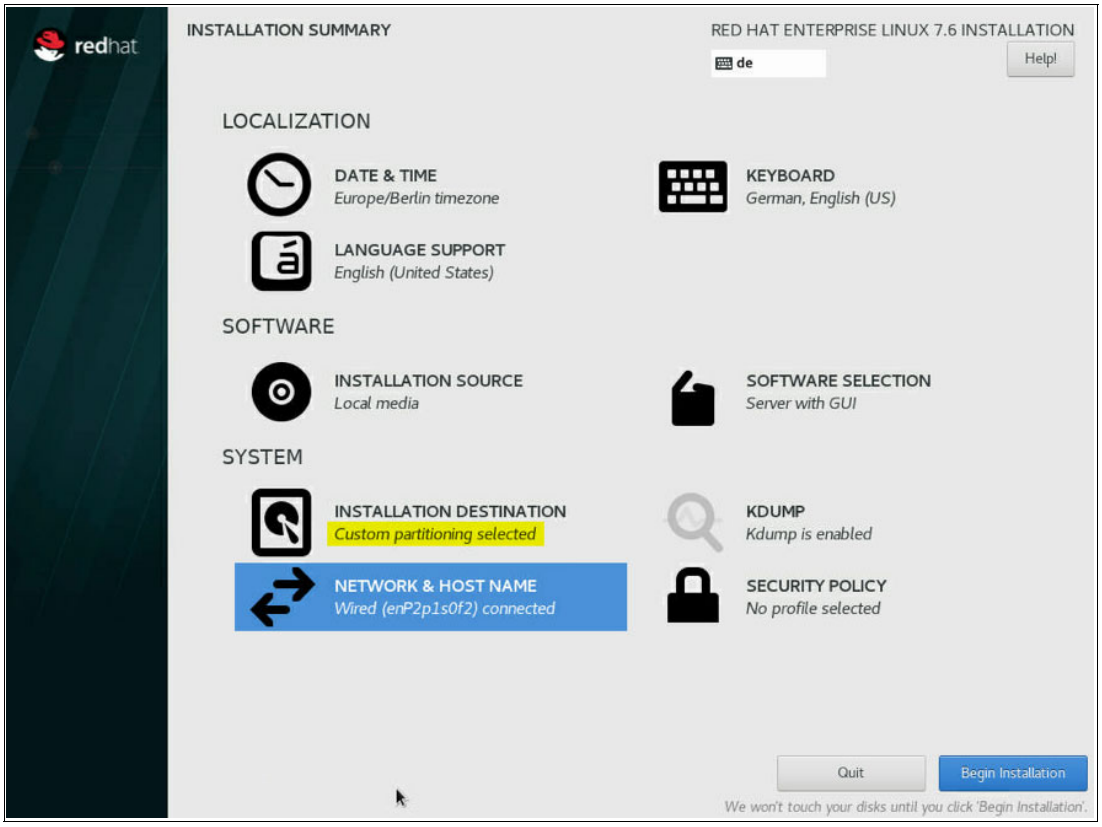


Figure 4-32 Red Hat Enterprise Linux 7.6 Installation window: INSTALLATION SUMMARY

4.4 Red Hat Virtualization

RHV is an x86 virtualization product that is produced by Red Hat and based on the KVM hypervisor.

This section describes the integration of an RHEL server installation into a KVM virtualization environment and an example of creating a Linux virtual machine (VM).

4.4.1 Configuring the host OS

To register the host OS in the Satellite Server, the correct time and date are necessary.

Tip: When configuring the ntp-server during installation, for example, `0.rhel.pool.ntp.org`, `1.rhel.pool.ntp.org`, `2.rhel.pool.ntp.org`, and `3.rhel.pool.ntp.org`, if the ntp-server is not configured, run the following command:

```
# ntpdate 0.rhel.pool.ntp.org
```

Enabling the Red Hat Enterprise Linux host repositories

To use an RHEL machine as a host, you must register the system with the Content Delivery Network, attach the RHEL server and RHV subscriptions, and enable the host repositories by completing the following steps[2]:

1. Register your system with the Content Delivery Network by running the following command, and enter your Customer Portal user name and password when prompted:

   ```
   # subscription-manager register
   ```

2. Find the RHEL server and RHV subscription pools and record the pool IDs by running the following command:

   ```
   # subscription-manager list –available
   ```

3. Use the pool IDs to attach the subscriptions to the system by running the following command:

   ```
   # subscription-manager attach --pool=poolid
   ```

   Note: To view currently attached subscriptions, run the following command:

   ```
   # subscription-manager list --consumed
   ```

   To list all enabled repositories (Figure 4-33), run the following command:

   ```
   # yum repolist
   ```

4. Configure the repositories as follows:

   ```
   # subscription-manager repos \
      --disable='*' \
      --enable=rhel-7-server-rpms \
      --enable=rhel-7-server-rhv-4-mgmt-agent-rpms \
      --enable=rhel-7-server-ansible-2-rpms
   ```

   For RHEL 7 hosts Little Endian on IBM POWER8 processor-based hardware, configure the repositories as follows:

   ```
   # subscription-manager repos \
       --disable='*' \
       --enable=rhel-7-server-rhv-4-mgmt-agent-for-power-le-rpms \
       --enable=rhel-7-for-power-le-rpms
   ```

   For RHEL 7 hosts Little Endian on IBM POWER9 processor-based hardware, configure the repositories as follows:

   ```
   # subscription-manager repos \
   --disable='*' \
       --enable=rhel-7-server-rhv-4-mgmt-agent-for-power-9-rpms \
       --enable=rhel-7-for-power-9-rpms
   ```

5. Ensure that all packages that are installed are up to date by running the following command:

   ```
   # yum update
   ```

[2] Source: https://red.ht/2EFrbLk

6. Restart the machine to see the list of enabled repositories.

```
[root@is334kvm: ~]# yum repolist
Loaded plugins: langpacks, product-id, search-disabled-repos, subscription-manager, vdsmupgrade
repo id                                                repo name                                                                                    status
rhel-7-for-power-9-rpms/7Server/ppc64le                Red Hat Enterprise Linux 7 for POWER9 (RPMs)                                                  8,519
rhel-7-for-power-le-rpms/7Server/ppc64le               Red Hat Enterprise Linux 7 for IBM Power LE (RPMs)                                           17,734
rhel-7-server-rhv-4-mgmt-agent-for-power-9-rpms/ppc64le Red Hat Virtualization 4 Management Agents (for RHEL 7 Server for IBM POWER9) RPMs               523
repolist: 26,776
[root@is334kvm: ~]#
```

Figure 4-33 Listing the enabled repositories from the CLI

You can also list the available repositories by using the Subscription Manager GUI (Figure 4-34). To open the Subscription Manager GUI, run the following command:

```
# subscription-manager-gui
```

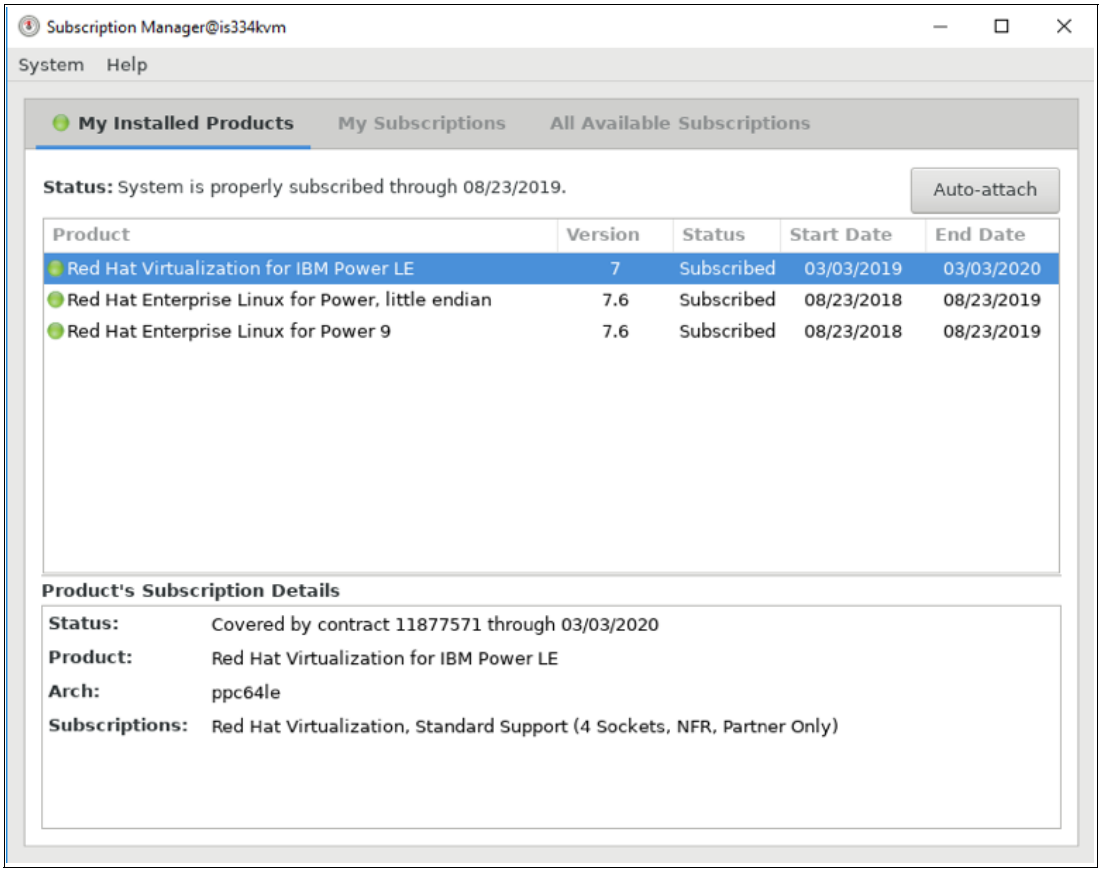


Figure 4-34 Subscription Manager GUI: Installed products

4.4.2 Red Hat Virtualization Manager

The RHV-M is the web GUI for the RHV software. All administration and operation tasks on compute nodes, network, and storage can be done by using this GUI.

Creating an RHV cluster

Before creating a new cluster, ensure that there is at least one host that is available to be assigned to it. All hosts in a cluster must have CPUs belonging to the same family as the cluster[3].

[3] Source: https://red.ht/2Q2RvUW

A data center can contain multiple clusters, and a cluster can contain multiple hosts. All hosts in a cluster must have the same CPU type (Intel or AMD). As a best practice, create your hosts before you create your cluster to ensure CPU type optimization. However, you can configure the hosts later by clicking **Guide Me**.

To create a cluster, complete the following steps:

1. Select **Compute** → **Clusters**, as shown in Figure 4-35.
2. Click **New**.
3. Select the **Data Center** to which the cluster will belong to from the drop-down list.
4. Enter the **Name** and **Description** of the cluster.
5. Select a network from the **Management Network** drop-down list to assign the management network role.
6. Select the **CPU Architecture** and **CPU Type** from the drop-down lists. It is important to match the CPU processor family with the minimum CPU processor type of the hosts you intend to attach to the cluster, or the host will be non-operational.
7. Select the **Compatibility Version** of the cluster from the drop-down list.

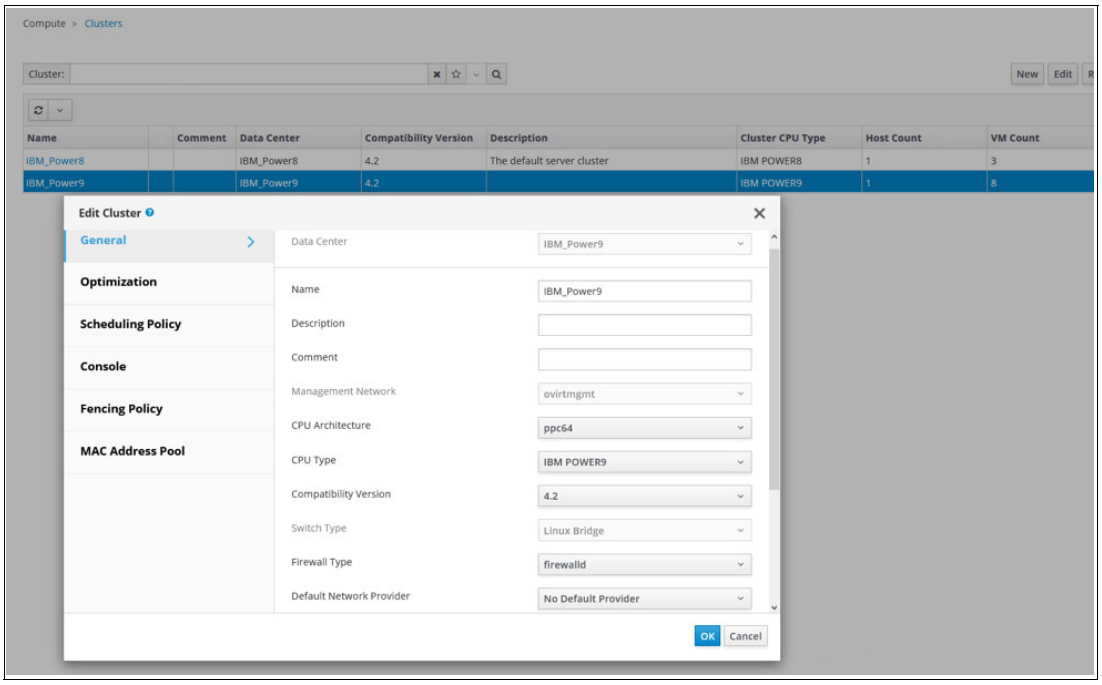


Figure 4-35 Compute Clusters window: Edit Cluster

Adding a host OS to an RHV cluster

Adding a host to your RHV cluster can take some time because the following steps are completed by the platform: virtualization checks, installation of packages, creation of a bridge, and a restart of the host. Use the details window to monitor the process as the host and the RHV-M establish a connection[4].

1. Select **Compute** → **Hosts**.
2. Click **New**.
3. Use the drop-down list to select the **Data Center** and **Host Cluster** for the new host.

[4] Source: https://red.ht/34Jb9dK

4. Enter the **Name** and the **Address** of the new host. The standard SSH port, port 22, is auto-filled in the **SSH Port** field.
5. Select an authentication method to use for the Manager to access the host:
   - Enter the root user's password to use password authentication.
   - Alternatively, copy the key that is shown in the **SSH PublicKey** field to `/root/.ssh/authorized_keys` on the host to use public key authentication.
6. Click **Advanced Parameters** to expand the advanced host settings:
   - Optionally, disable the automatic firewall configuration.
   - Optionally, add a host SSH fingerprint to increase security. You can add it manually or let it be fetched automatically.
7. Optionally, configure power management, where the host has a supported power management card. For information about power management configuration, see 7.5.4.2 Host Power Management Settings Explained.
8. Click **OK**, as shown in Figure 4-36.

The new host appears in the list of hosts with a status of Installing, and you can view the progress of the installation in the details window. After a brief delay, the host status changes to Up.

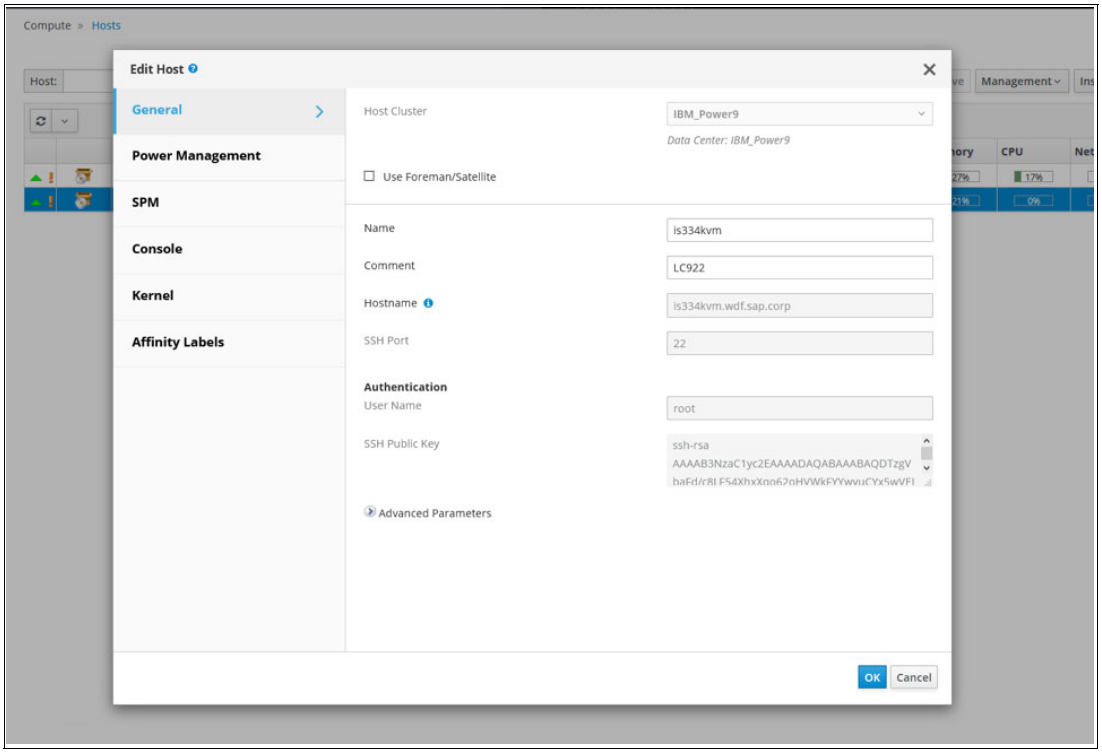


Figure 4-36 Compute Host window: Edit Host

4.4.3 Guest OS deployment and optimization

This section describes how to deploy a new VM. Complete the following steps:

1. Select **Compute** → **Virtual Machines**, as shown in Figure 4-37.

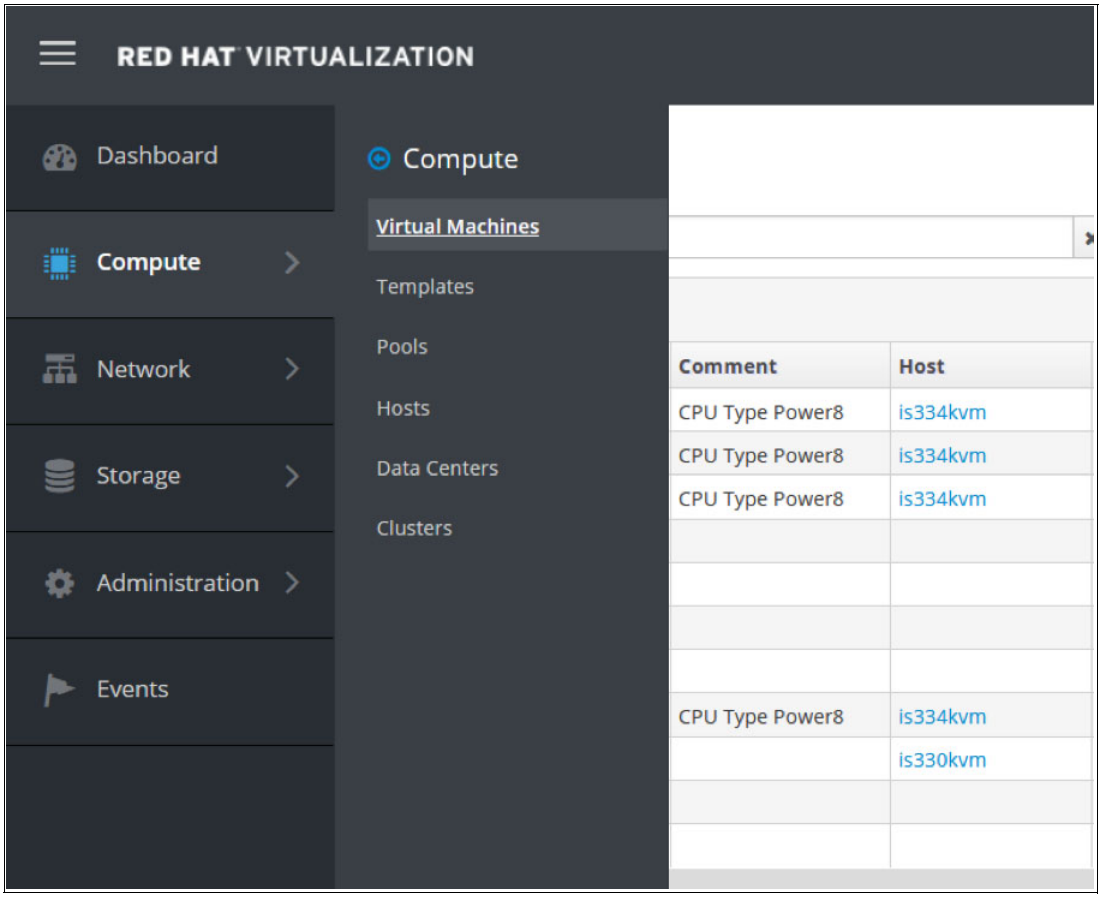


Figure 4-37 Red Hat Virtualization window

2. Click **New**.

3. Complete the fields in the General menu, as shown in Figure 4-38.

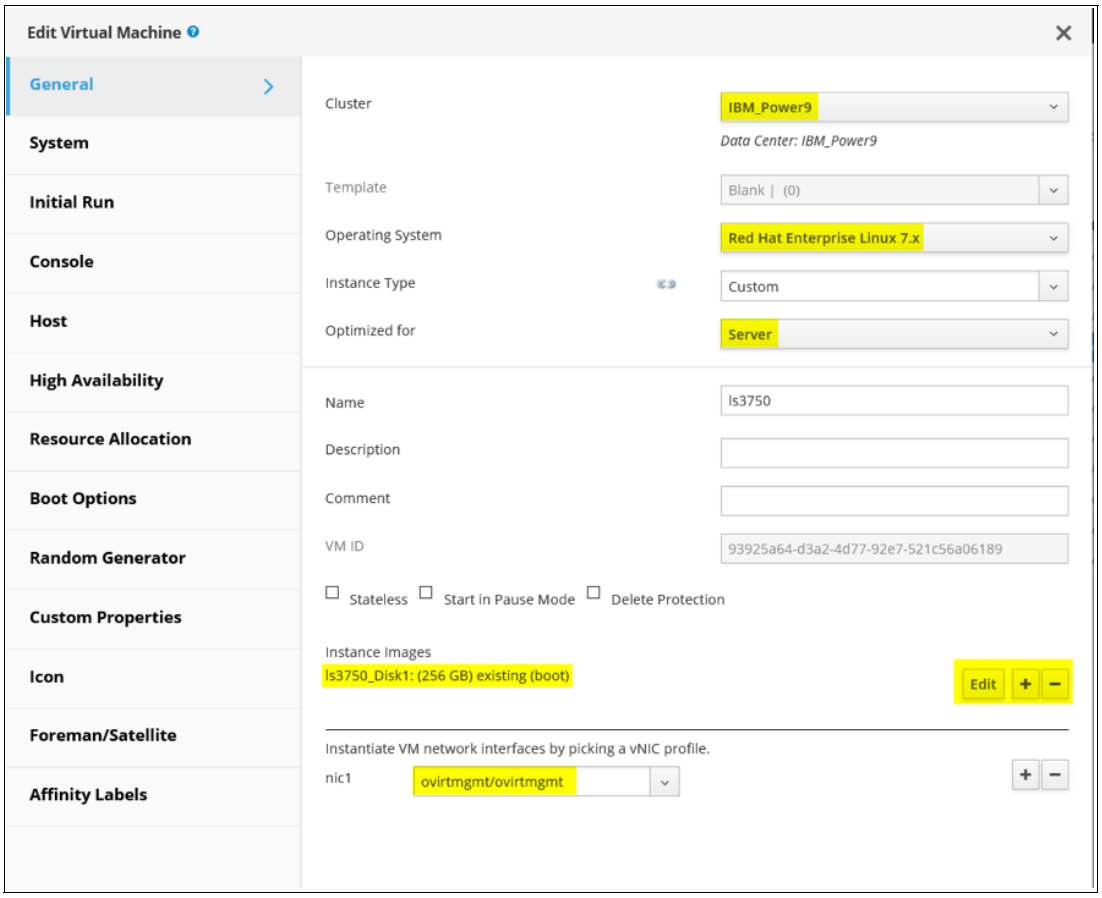


Figure 4-38 Edit Virtual Machine window: General

4. Complete the fields in the System menu, as shown in Figure 4-39.

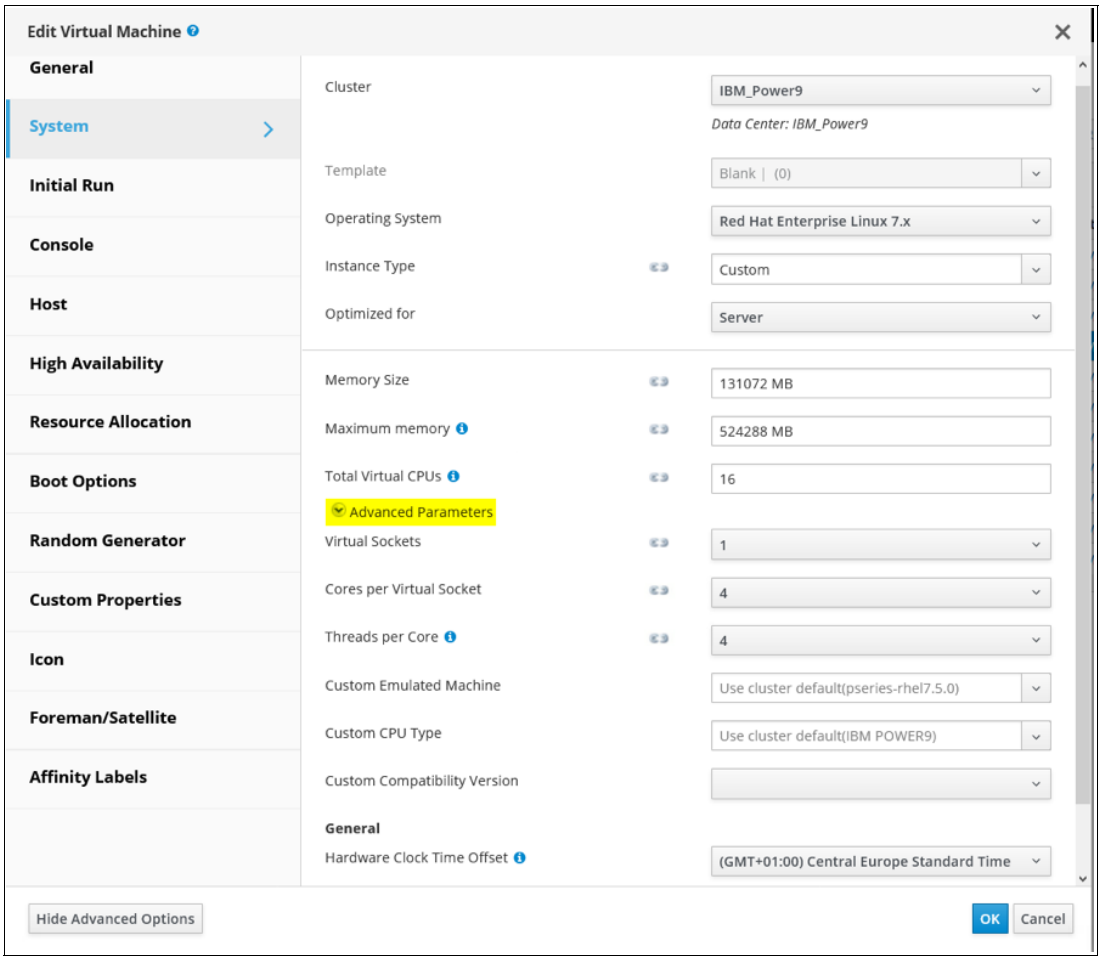


Figure 4-39 Edit Virtual Machine window: System

5. Complete the fields in the Console menu, as shown in Figure 4-40.

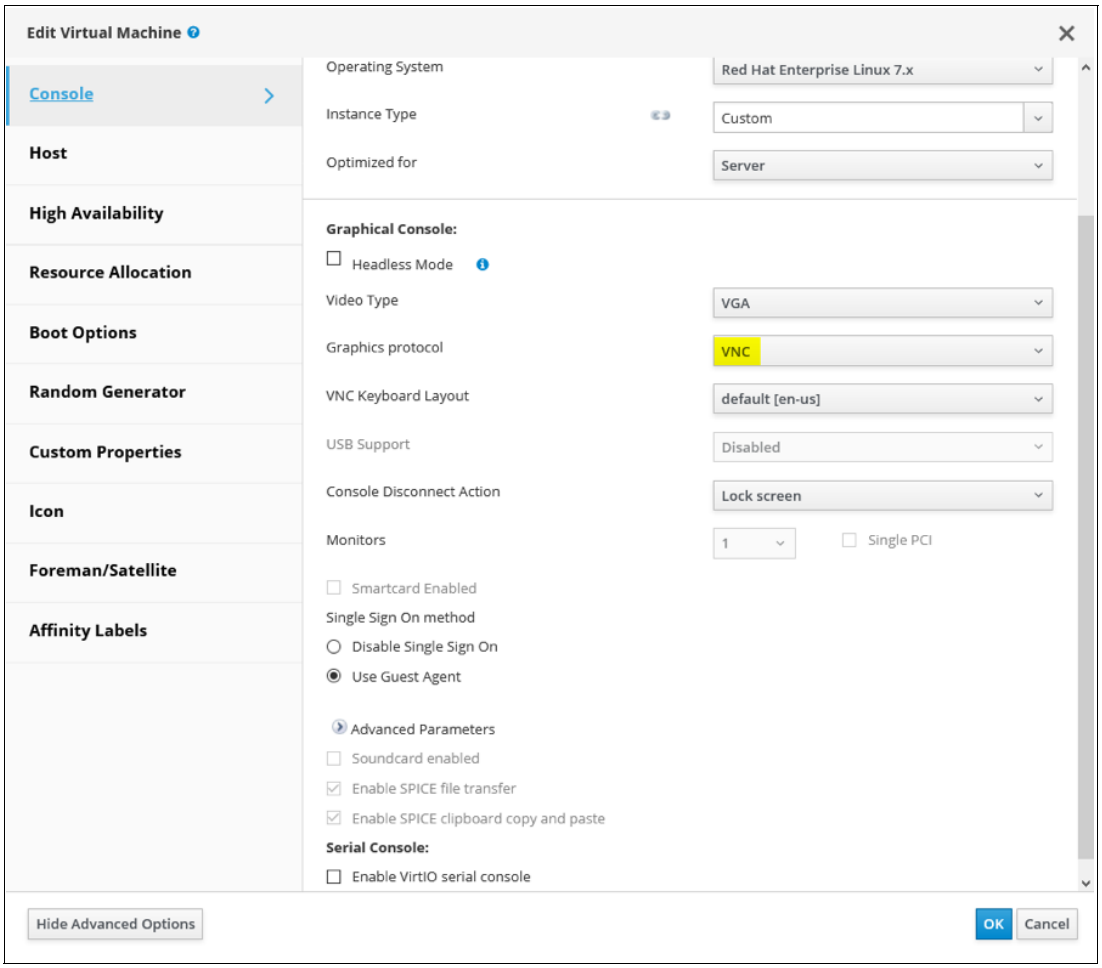


Figure 4-40 Edit Virtual Machine window: Console

6. Complete the fields in the Host menu, as shown in Figure 4-41.

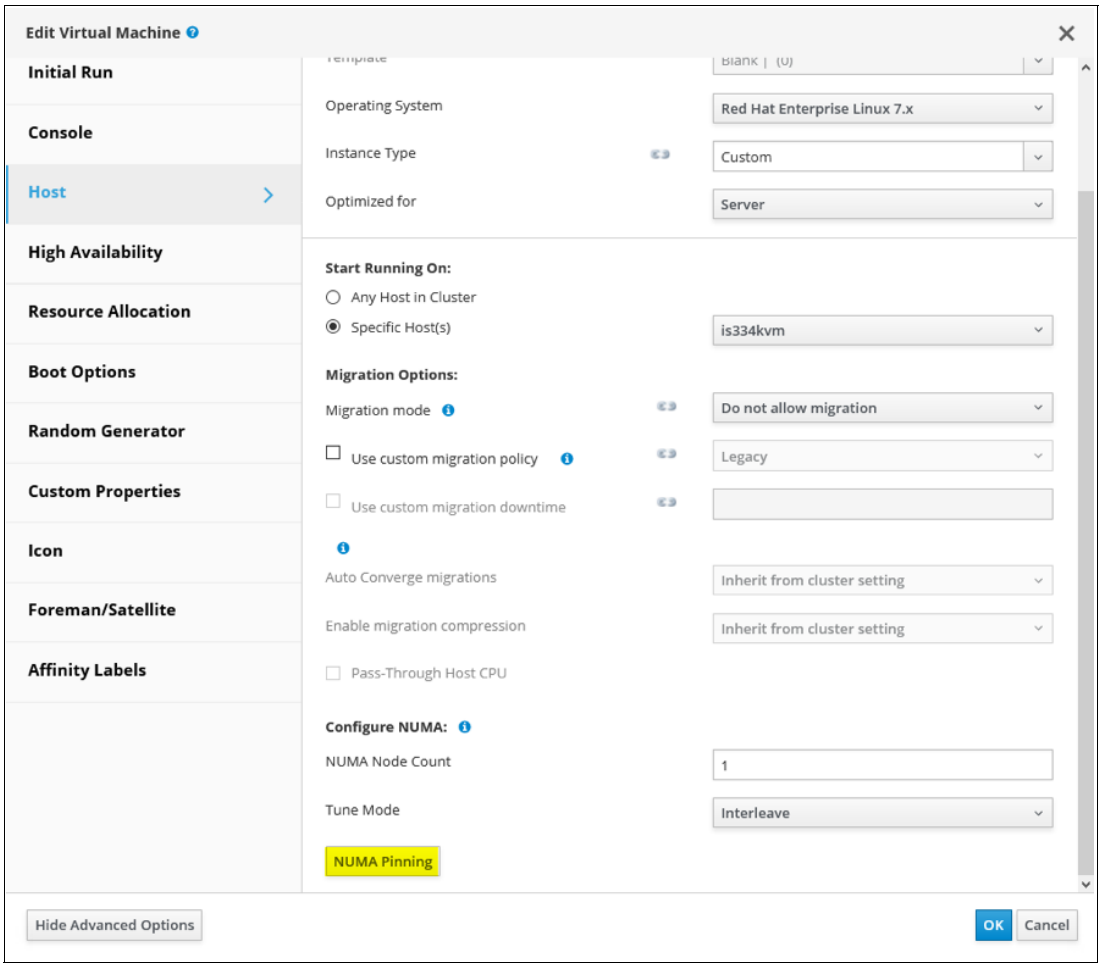


Figure 4-41 Edit Virtual Machine window: Host

7. Configure the non-uniform memory access (NUMA) topology, as shown in Figure 4-42. Performance optimization is used to pin VMs to dedicated bare metal CPU sockets (NUMA Pinning and NUMA Topology).

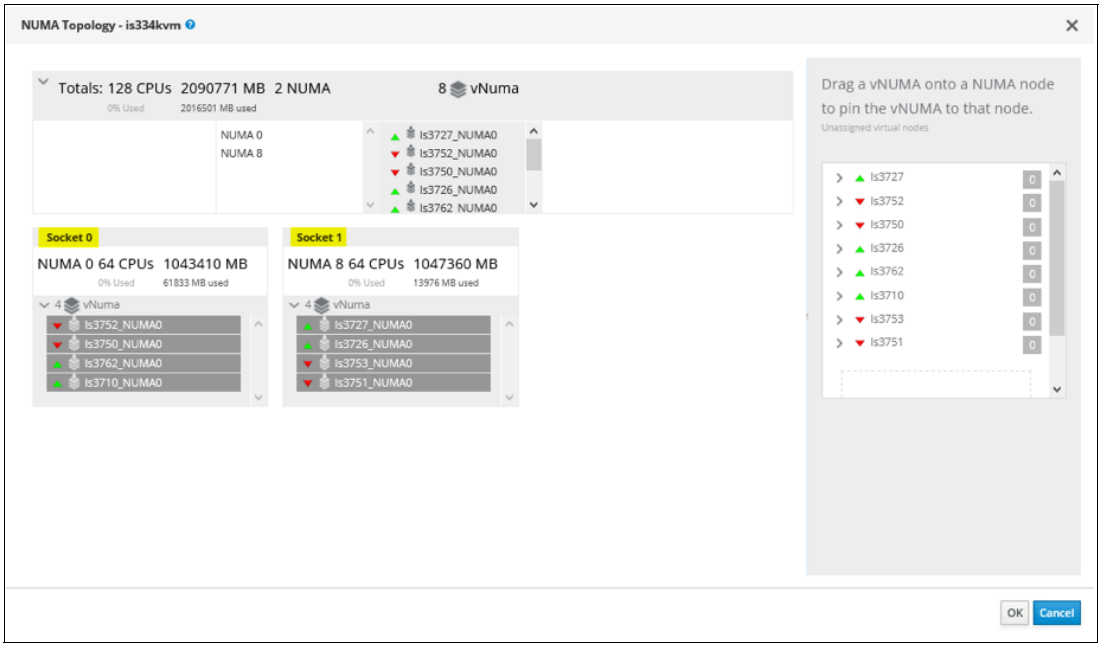


Figure 4-42 NUMA Topology window

8. Complete the fields in the Resource Allocation window for the VM, as shown in Figure 4-43.

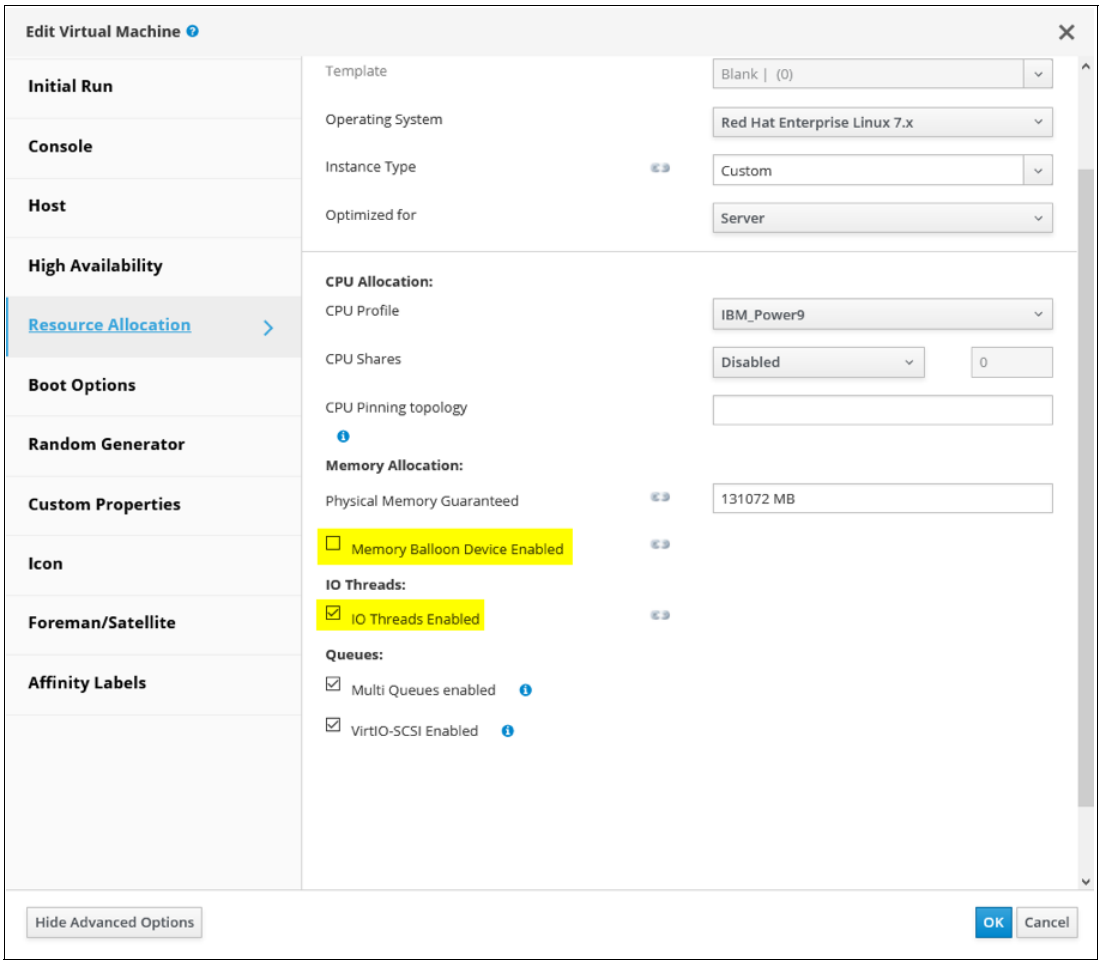


Figure 4-43 Edit Virtual Machine window: Resource Allocation

9. Complete the fields in the Boot Options window, as shown in Figure 4-44.

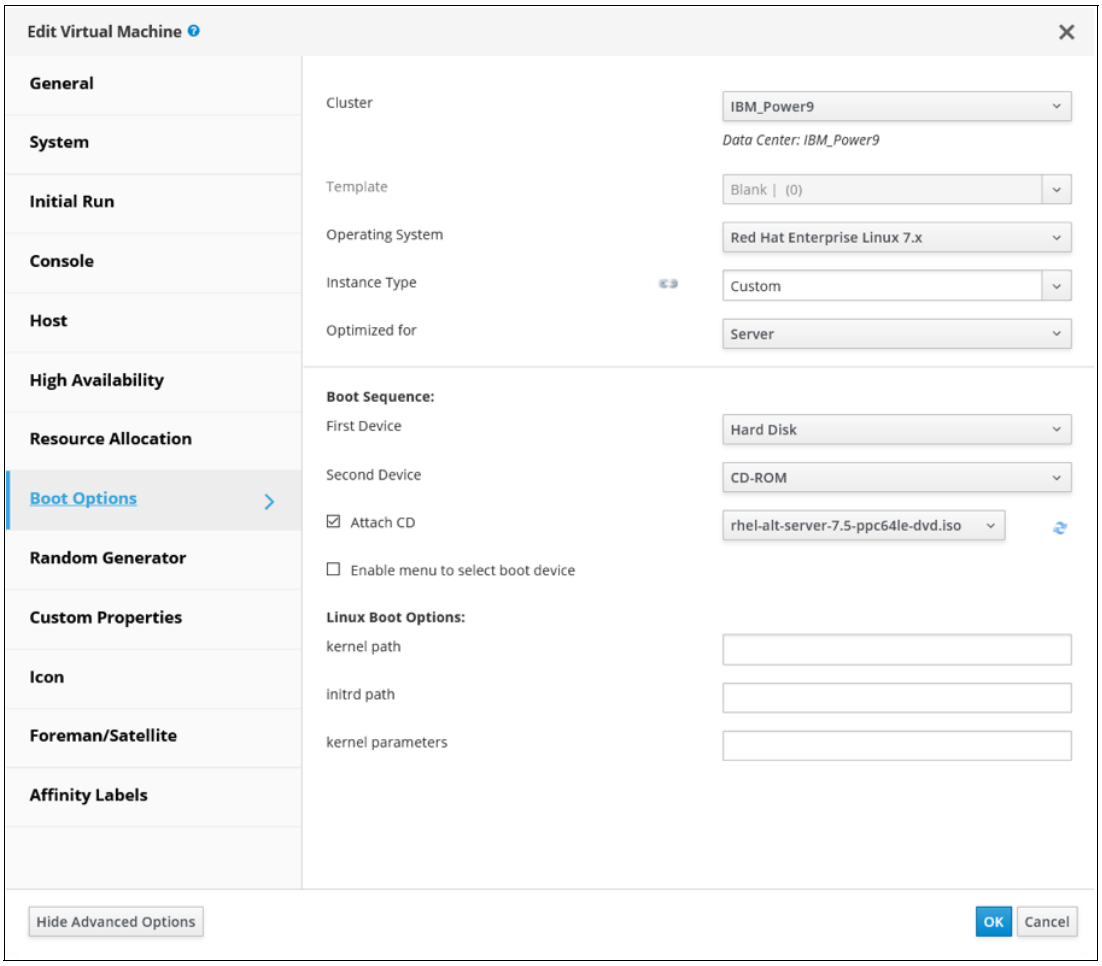


Figure 4-44 Edit Virtual Machine window: Boot Options

10. Complete the fields in the Custom Properties window, as described in “Setup instructions” on page 99.
11. Click **OK**.

4.4.4 Guest OS monitoring

This section shows how to monitor a guest OS.

RHV Manager monitoring features

Basic usage of memory, CPU, and the network is shown in real time on the Virtual Machines overview window, as shown in Figure 4-45.

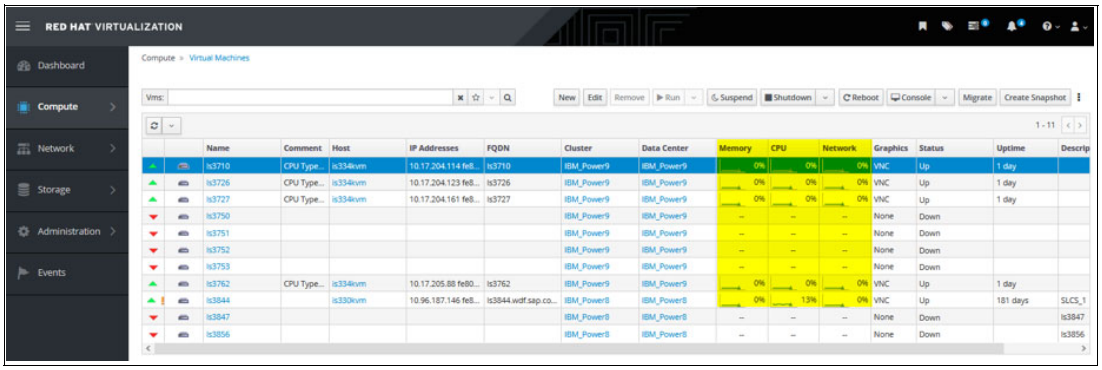


Figure 4-45 Red Hat Virtualization: VM monitoring

The nmon tool for Linux

Nigel's performance Monitor for Linux on Power, x86, x86_64, Mainframe & ARM (Raspberry Pi) (**nmon**) provides a vast amount of important performance information in a simple terminal session, as shown in Figure 4-46. It can output the data in two ways:

- Real time.
- Save the data to a comma-separated file for analysis and longer term data capture.

```
nmon-16f──────────────────Hostname=is334kvm─────Refresh= 2secs ──16:49.56──────────

 ----------------------------
 _ __  _ __ ___   ___  _ __
| '_ \| '_ ` _ \ / _ \| '_ \
| | | | | | | | | (_) | | | |
|_| |_|_| |_| |_|\___/|_| |_|
 ----------------------------
                                For help type H or ...
                                 nmon -?  - hint
                                 nmon -h  - full details

                                To stop nmon type q to Quit

NAME="Red Hat Enterprise Linux Server"
Native POWER9, altivec supported
Native owns all 128 CPUs & SMT=4
Native firmware : OPAL
Processor Clock=3800.000000MHz            Little Endian

Use these keys to toggle statistics on/off:
  c = CPU         l = CPU Long-term     - = Faster screen updates
  C = " WideView  U = Utilisation       + = Slower screen updates
  m = Memory      V = Virtual memory    j = File Systems
  d = Disks       n = Network           . = only busy disks/procs
  r = Resource    N = NFS               h = more options
  k = Kernel      t = Top-processes     q = Quit
```

Figure 4-46 The nmon performance monitoring pane

The tool can be downloaded from nmon for Linux or from Tools for Linux on Power - Service and productivity tools.

To start an SSH shell to interact with **nmon**, run the following command:

```
[root@is334kvm: ~]# nmon
```

The interactive Mode Keys in Alphabetical Order are as follows:

```
Key --- Toggles on off to control what is displayed ---
b   = Black and white mode (or use -b command line option)
c   = CPU Utilisation stats with bar graphs (CPU core threads)
C   = CPU Utilisation as above but concise wide view (up to 192 CPUs)
d   = Disk I/O Busy% & Graphs of Read and Write KB/s
D   = Disk I/O Numbers including Transfers, Average Block Size & Peaks
      (type: 0 to reset)
g   = User Defined Disk Groups
      (assumes -g <file> when starting nmon)
G   = Change Disk stats (d) to just disks
      (assumes -g auto   when starting nmon)
h   = This help information
j   = File Systems including Journal File Systems
k   = Kernel stats Run Queue, context-switch, fork, Load Average & Uptime
l   = Long term Total CPU (over 75 snapshots) via bar graphs
L   = Large and =Huge memory page stats
m   = Memory & Swap stats
M   = MHz for machines with variable frequency 1st=Threads 2nd=Cores 3=Graphs
n   = Network stats & errors (if no errors it disappears)
N   = NFS - Network File System
      1st NFS V2 & V3, 2nd=NFS4-Client & 3rd=NFS4-Server
o   = Disk I/O Map (one character per disk pixels showing how busy it is)
      Particularly good if you have 100's of disks
p   = PowerVM LPAR Stats from /proc/ppc64/lparcfg
q   = Quit
r   = Resources: Machine type, name, cache details &
      OS version & Distro + LPAR
t   = Top Processes: select the data & order 1=Basic, 3=Perf,
      4=Size, 5=I/O=root only
u   = Top Process with command line details
U   = CPU utilisation stats - all 10 Linux stats:
      user, user_nice, system, idle, iowait, irq, softirq, steal,
      guest, guest_nice
v   = Experimental Verbose mode - tries to make recommendations
V   = Virtual Memory stats
```

To start the **nmon** daemon to log long-term CPU utilization data collection, use the following command syntax:

```
nmon -f [-s <seconds>] [-c <count>] [-T] [-m <directory>]
-f spreadsheet output format [note: default -s300 -c288]
   output file is <hostname>_YYYYMMDD_HHMM.nmon
-s <seconds>  Time between snap shots -
              with "-c count" decides duration of the data capture
-c <number>   The number of snapshots before nmon stops
-T  Include Top Processes in the output and
    saves command line arguments in UARG section
-m <directory> nmon changes to this directory before saving to file
```

For example,

```
[root@is334kvm: ~]# mkdir -p /tmp/nmon4html ; ~/nmon/nmon -f -T -s 20 -c 30 -m
/tmp/nmon4html
```

The nmonchart tool: Web page graphs from nmon data

After the nmon data collection is complete, the collected data can be converted into HTML files by using the **nmonchart** tool, as shown in Figure 4-47.

The command syntax is:

```
nmonchart <nmon-file> <output-file>.html
```

For example:

```
nmonchart is330kvm_180807_1018.nmon is330kvm_180807_1018.html
```

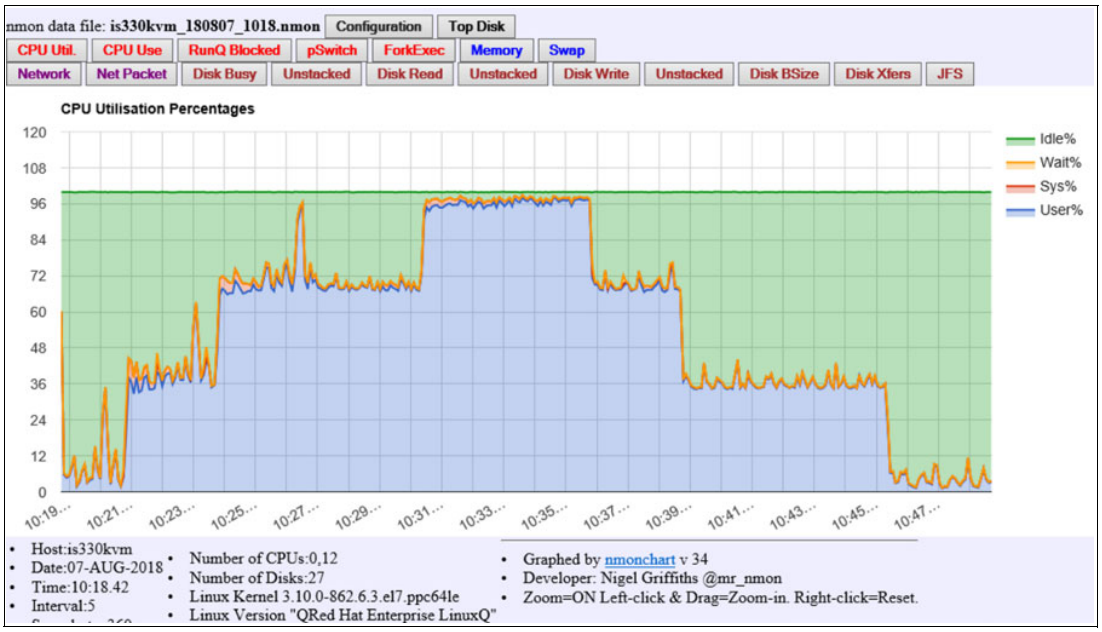


Figure 4-47 A nmonchart graph

4.4.5 SAP monitoring

This section provides details about SAP monitoring.

Host OS and SAP monitoring enablement

The SAP monitoring software (**saposcol**), which is installed and running on each KVM guest along with the SAP solution stack, requires access to platform metrics data that is available only on the KVM host side. To establish a channel between the host and the guest sides that facilitate the data transfer of these matrixes, the Virtual Host Metrics Daemon (**vhostmd**) is used. The **vhostmd** daemon itself uses **libvirt** to collect the required host side metrics. For more information, see SAP Note 2097317.

Configuring SAP Monitoring by using vhostmd in RHV

Machine monitoring for SAP applications uses the **vhostmd vdsm** hook, which is provided to hosts by the `vdsm-hook-vhostmd` package. Red Hat Virtualization Hypervisor (RHVH) hosts include the `vdsm-hook-vhostmd` package by default. However, the `vdsm-hook-vhostmd` package must be installed on RHEL hosts.

For more information, see How to configure SAP Monitoring by using vhostmd in RHV?

Setup instructions

The following section describes how to set up the `vdsm-hook-vhostmd` package on the RHEL host, enable SAP monitoring for a VM, and verify the monitoring:

1. As root, install the `vdsm-hook-vhostmd` package on the RHEL host by running the following command:

   ```
   # yum install vdsm-hook-vhostmd
   ```

1. Select **Compute** → **Virtual Machines**.
2. Select **VM**.
3. Edit the VM by selecting **Custom Properties** and selecting **sap_agent**, as shown in Figure 4-48.

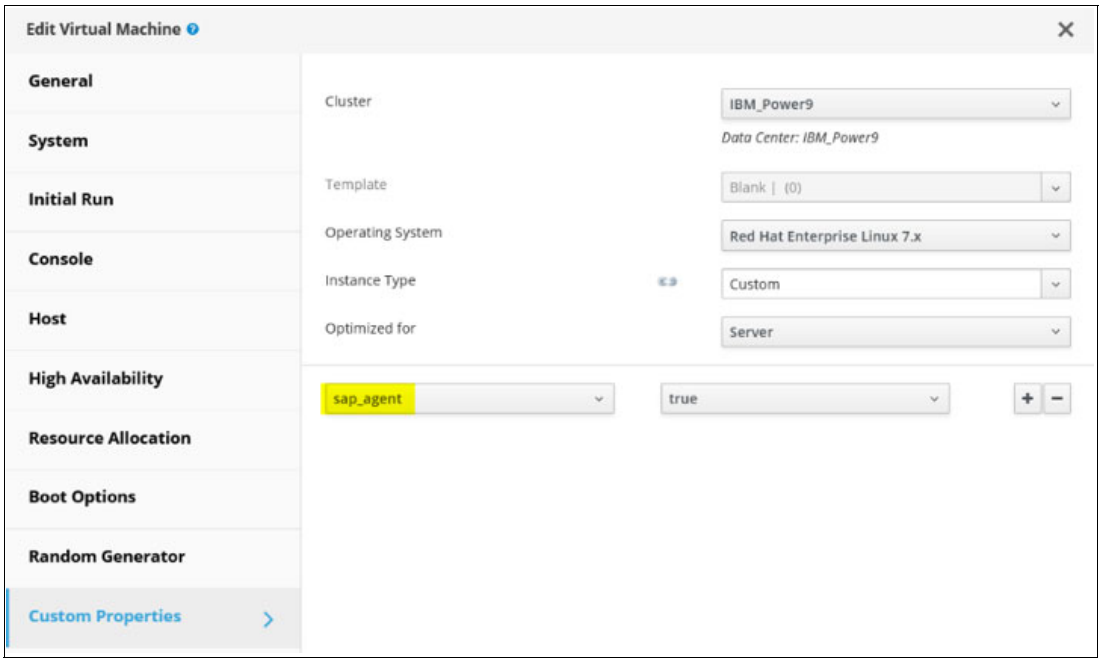


Figure 4-48 Edit Virtual Machine window: Configuring Custom Properties

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 paper.

IBM Redbooks

The following IBM Redbooks publications provide more information about the topics in this document. Some publications that are referenced in this list might be available in softcopy only.

- *IBM PowerKVM: Configuration and Use*, SG24-8231
- *IBM Power System E950: Technical Overview and Introduction*, REDP-5509
- *IBM Power System E980: Technical Overview and Introduction*, REDP-5510
- *IBM Power Systems H922 and H924 Technical Overview and Introduction*, REDP-5498
- *IBM Power System L922 Technical Overview and Introduction*, REDP-5496
- *IBM Power Systems LC921 and LC922: Technical Overview and Introduction*, REDP-5495
- *IBM Power Systems S922, S914, and S924 Technical Overview and Introduction*, REDP-5497
- *Live Partition Mobility Setup Checklist*, TIPS1184

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:

- The nmon home page

  http://nmon.sourceforge.net
- OpenPOWER Abstraction Layer (OPAL)

  https://github.com/open-power
- Quick Start Guide for Installing Linux on Linux Cloud and Cluster (LC) servers

  https://www.ibm.com/support/knowledgecenter/linuxonibm/liaaa/ic-homepage.htm
- Red Hat Virtualization (RHV)

  https://bit.ly/2tckA8J

Help from IBM

IBM Support and downloads

ibm.com/support

IBM Global Services

ibm.com/services

REDP-5579-00

ISBN 0738458503

Printed in U.S.A.