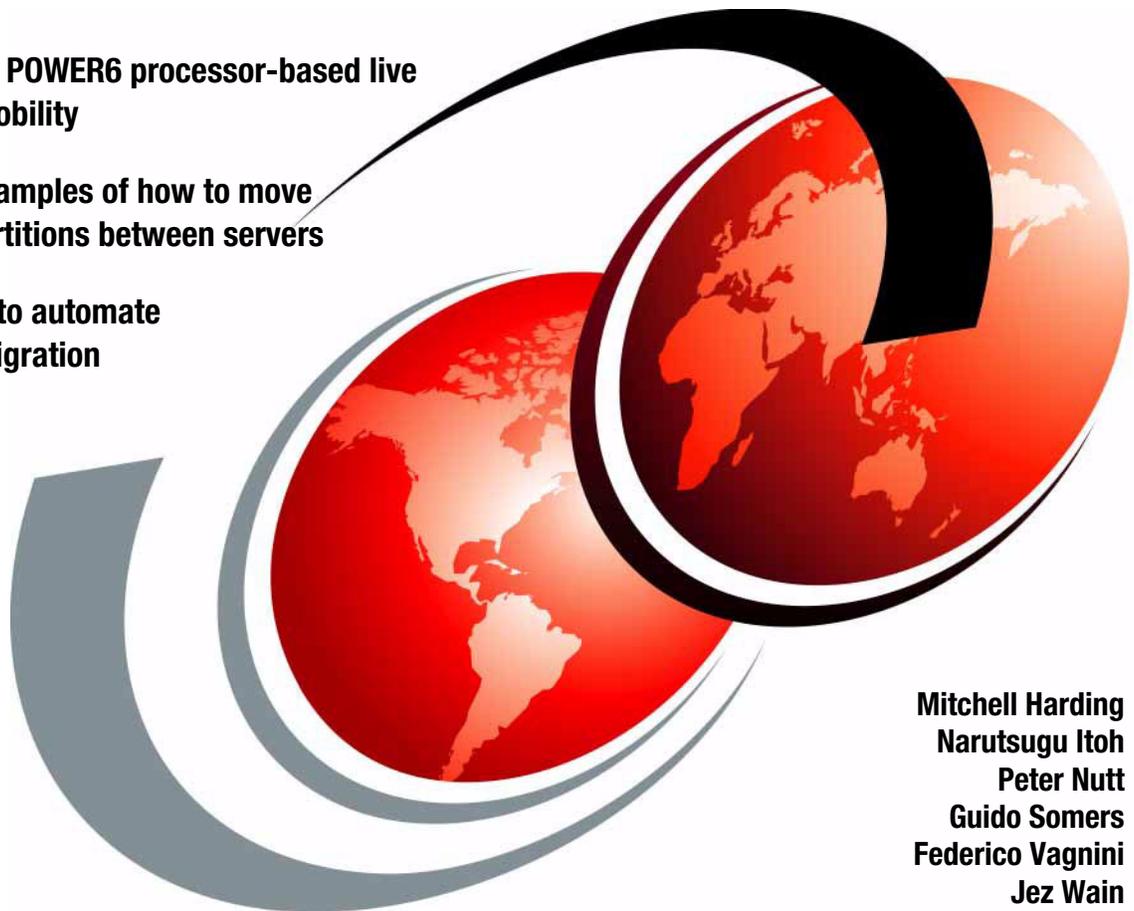**IBM**

IBM System p
Live Partition Mobility

Explore the POWER6 processor-based live partition mobility

Perform examples of how to move running partitions between servers

Learn how to automate partition migration

Mitchell Harding
Narutsugu Itoh
Peter Nutt
Guido Somers
Federico Vagnini
Jez Wain

Redbooks

IBM

International Technical Support Organization

IBM System p Live Partition Mobility

October 2007

First Edition (October 2007)

This edition applies to AIX Version 6.0, AIX 5L Version 5.3 TL7, HMC Version 7.3.2, and POWER6 based servers, such as the IBM System p 570 (9117-MMA).

Note: This book is based on a pre-GA version of a product and may not apply when the product becomes generally available. We recommend that you consult the product documentation or follow-on versions of this IBM Redbooks publication for more current information.

Contents

Figures

Tables

Notices

This information was developed for products and services offered in the U.S.A.

IBM may not offer the products, services, or features discussed in this document in other countries. Consult your local IBM representative for information on the products and services currently available in your area. Any reference to an IBM product, program, or service is not intended to state or imply that only that IBM product, program, or service may be used. Any functionally equivalent product, program, or service that does not infringe any IBM intellectual property right may be used instead. However, it is the user's responsibility to evaluate and verify the operation of any non-IBM product, program, or service.

IBM may have patents or pending patent applications covering subject matter described in this document. The furnishing of this document does not give you any license to these patents. You can send license inquiries, in writing, to:
IBM Director of Licensing, IBM Corporation, North Castle Drive, Armonk, NY 10504-1785 U.S.A.

The following paragraph does not apply to the United Kingdom or any other country where such provisions are inconsistent with local law: INTERNATIONAL BUSINESS MACHINES CORPORATION PROVIDES THIS PUBLICATION "AS IS" WITHOUT WARRANTY OF ANY KIND, EITHER EXPRESS OR IMPLIED, INCLUDING, BUT NOT LIMITED TO, THE IMPLIED WARRANTIES OF NON-INFRINGEMENT, MERCHANTABILITY OR FITNESS FOR A PARTICULAR PURPOSE. Some states do not allow disclaimer of express or implied warranties in certain transactions, therefore, this statement may not apply to you.

This information could include technical inaccuracies or typographical errors. Changes are periodically made to the information herein; these changes will be incorporated in new editions of the publication. IBM may make improvements and/or changes in the product(s) and/or the program(s) described in this publication at any time without notice.

Any references in this information to non-IBM Web sites are provided for convenience only and do not in any manner serve as an endorsement of those Web sites. The materials at those Web sites are not part of the materials for this IBM product and use of those Web sites is at your own risk.

IBM may use or distribute any of the information you supply in any way it believes appropriate without incurring any obligation to you.

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

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

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

COPYRIGHT LICENSE:

This information contains sample application programs in source language, which illustrate programming

xiii

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.

Trademarks

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

AIX 5L™	HACMP™	PTX®
AIX®	i5/OS®	Redbooks®
DB2®	IBM®	Redbooks (logo) ®
Enterprise Storage Server®	Parallel Sysplex®	RS/6000®
General Parallel File System™	POWER™	System p™
Geographically Dispersed	POWER Hypervisor™	Tivoli®
Parallel Sysplex™	POWER4™	Workplace™
GDPS®	POWER5™	
GPFS™	POWER6™	

The following terms are trademarks of other companies:

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

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

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

Preface

Live Partition Mobility allows you to migrate running AIX® and Linux® partitions and their hosted applications from one physical server to another without disrupting infrastructure services. The migration transfers the entire partition state, including processor context, memory, attached virtual devices, and connected users.

Live Partition Mobility helps you meet increasingly stringent service-level agreements (SLAs) because it allows you to proactively move running partitions and applications from systems requiring intervention or under heavy load to other systems.

Live Partition Mobility can be automated and incorporated into system management tools and scripts. Support for multiple concurrent migrations allows you to rebalance the load across systems very quickly. For single partition, user initiated migrations, the HMC offers an easy-to-use migration wizard.

Live Partition Mobility is the next step in the IBMs System p™ virtualization continuum. It can be combined with other virtualization technologies, such as logical partitions, Live Workload Partitions, and the SAN Volume Controller, to provide a fully virtualized computing platform that offers the degree of system and infrastructure flexibility required by today's production data centers.

This IBM® Redbooks® publication assists you in understanding, planning, preparing for, and performing partition migration on IBM System p servers running AIX.

The team that wrote this IBM Redbooks publication

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

Mitchell Harding is a Staff Software Engineer working in the IBM System p test organization for IBM USA. He holds a degree in Computer Science from the University of Texas at Austin. His areas of interest include AIX, storage area networks, virtualization, and software testing.

Narutsugu Itoh is an Advisory IT Specialist working for IBM Japan, Technical Sales Support, Advanced Technical Support (ATS) supporting field sales. He holds a degree in Electrical Engineering from Waseda University. He has 12 years of experience with IBM System p servers. He supported the Nagano Winter Olympic Games in 1998 where RS/6000® and SP systems were used to host the official game records and Web servers. He is co-hosting a local residency program for Asia-Pacific System p experts. His areas of expertise include system implementation, logical partitioning, virtualization, AIX, performance tuning, and SAN configuration.

Peter Nutt is a Senior IT Specialist based in the UK and joined IBM in 2000. Peter primarily works on AIX systems and has expertise in the fields of virtualization, performance analysis, and implementation. Peter has co-authored two other IBM Redbooks and is a frequent instructor at IBM Technical Universities.

Guido Somers is a Senior Accredited IT Specialist working for IBM Belgium. He has 12 years of experience in the Information Technology field, nine years of which were within IBM. He holds degrees in Biotechnology, Business Administration, Chemistry, and Electronics, and did research in the field of Theoretical Physics. His areas of expertise include AIX, system performance and tuning, logical partitioning, virtualization, HACMP™, SAN, IBM System p servers, as well as other IBM hardware offerings. He currently works as a Client IT Architect for Infrastructure and Global ISV Solutions in the e-Business Solutions Technical Support (eTS) organization. He is the author of the second edition of *Integrated Virtualization Manager on IBM System p5*, REDP-4061.

Federico Vagnini is a Certified IT Specialist working in Field Technical Sales Support organization in IBM Italy, supporting IBM sales, Business Partners, and clients with pre-sales consultation and implementation on high-end system environments. He has 12 years of experience in IBM System p servers and he is a member of the Virtualization Technology Focus Group. He is the author of Redbooks about AIX and Integrated Virtualization Manager.

Jez Wain is an infrastructure architect working for Bull in Grenoble, France, applying today's virtualization technologies to end-to-end solutions. He has worked on real-time airborne embedded systems, fly-by-wire control, telemetry and data-acquisition, AIX kernel, and Java™ middleware. He has contributed to other IBM publications.

The project that produced this document was managed by:

Scott Vetter, PMP

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

Become a published author

Join us for a two- to six-week residency program! Help write an IBM Redbooks publication dealing with specific products or solutions, while getting hands-on experience with leading-edge technologies. You will have the opportunity to team with IBM technical professionals, Business Partners, and Clients.

Your efforts will help increase product acceptance and customer satisfaction. As a bonus, you will develop a network of contacts in IBM development labs, and increase your productivity and marketability.

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 Redbooks to be as helpful as possible. Send us your comments about this book or other Redbooks in one of the following ways:

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

   ibm.com/redbooks

► Send your comments in an e-mail to:

   redbooks@us.ibm.com

► Mail your comments to:

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

Overview

Live Partition Mobility allows you to migrate running AIX and Linux partitions and their hosted applications from one physical server to another without disrupting the infrastructure services. The migration operation, which takes just a few seconds, maintains complete system transactional integrity. The migration transfers the entire system environment, including processor state, memory, attached virtual devices, and connected users.

IBM System p systems are designed to offer the highest stand-alone availability in the industry. Enterprises must occasionally restructure their infrastructure to meet new IT requirements. By letting you move your running production applications from one physical server to another, Live Partition Mobility allows for non-disruptive maintenance or modification to a system (to your users). This mitigates the impact on partitions and applications formerly caused by the occasional need to shut down a system.

Today, even small IBM System p servers frequently host many logical partitions. As the number of hosted partitions increases, finding a maintenance window acceptable to all becomes increasingly difficult. Live Partition Mobility allows you to move your partitions around such that you can perform previously disruptive operations on the machine when it best suits you, rather than when it causes the least inconvenience to the users.

Live Partition Mobility helps you meet continuously increasing stringent service-level agreements (SLAs) because it allows you to proactively move running partitions and applications from one server to another.

The ability to move running partitions from one server to another offers you the ability to balance workloads and resources. Should a key application's resource requirements peak unexpectedly to a point where there is contention for server resources, you might move it to a larger server or move other, less critical, partitions to different servers, and use the freed up resources to absorb the peak.

Live Partition Mobility may also be used as a mechanism for server consolidation, as it provides an easy path to move applications from individual, stand-alone servers to consolidation servers. If you have partitions with workloads that have widely fluctuating resource requirements over time (for example, with a peak workload at the end of the month or the end of the quarter) you can use Live Partition Mobility to consolidate partitions to a single server during the off-peak period, allowing you to power-off unused servers. Then move the partitions to their own, adequately configured servers, just prior to the peak. This approach also offers energy savings by reducing the power to run machines and the power to keep them cool during off-peak periods.

Live Partition Mobility can be automated and incorporated into system management tools and scripts. Support for multiple concurrent migrations allows you to liberate systems very quickly. For single-partition, point-in-time migrations, the HMC offers an easy-to-use migration wizard.

Live Partition Mobility contributes to the continuous availability goal. It can:

► Reduce planned down time by dynamically moving applications from one server to another.

► Respond to changing workloads and business requirements by letting you move workloads from heavily loaded servers to servers that have spare capacity.

► Reduce energy consumption by allowing you to easily consolidate workloads and power off unused servers.

Live Partition Mobility is the next step in the IBM System p virtualization continuum. It can be combined with other virtualization technologies, such as logical partitions, Live Workload Partitions, and the SAN Volume Controller to provide a fully virtualized computing platform offering the degree of system and infrastructure flexibility required by today's production data centers.

1.1 Introduction

A partition migration operation can occur either when a partition is powered off (inactive), or when a partition is providing service (active).

During an active partition migration, there is no disruption of system operation or user service. For example, a partition hosting a live production database with normal user activities can be migrated to a second system with no loss of data, no loss of connectivity, and no effect on running transactions.

A logical partition may be migrated between two POWER6™ based systems, provided the destination system has enough resources to host the partition. There is no restriction on processing units or partition memory size, either for inactive or for active migration.

In this chapter, we provide an overview of Live Partition Mobility with a high-level description of its features.

1.2 Cross-system flexibility is the requirement

Infrastructure flexibility has become a key criteria when designing and deploying information technology solutions. Application requirements frequently change and the hardware infrastructure they rely upon must be capable of adapting to the new requirements in a very short time, but also with minimal to no impact on the service level. Configuration changes must be applied in a very simple and secure way, with limited administrator intervention, to reduce change management costs and the related risk.

The Advanced POWER™ Virtualization feature introduced in POWER5™ based systems provides excellent flexibility capabilities within each system. The virtualization of processor capacity and the fine distribution of memory, combined with network and disk virtualization, enable administrators to create multiple fine-grained logical partitions within a single system. Computing power can be distributed among partitions automatically in real time, depending on real application needs, with no user action. System configuration changes are made by policy-based controls or by administrators with very simple and secure operations that do not interrupt service.

While single-system virtualization greatly improves the flexibility of an IT solution, customers' service requirements often demand a more comprehensive view of the entire infrastructure. In many instances, applications are distributed across multiple systems ensuring isolation, optimization of global system resources, and adaptability of the infrastructure to new workloads.

One of the most time consuming activities in a complex environment is the transfer of a workload from one system to another. There may be many reasons for the migration, for example:

► Resource balancing

   A system does not have enough resources for the workload while another system does

► New system deployment

   A workload running on an existing system must be migrated to a new, more powerful one.

► Availability requirements

   When a system requires maintenance, its hosted applications must not be stopped and can be migrated to another system.

Without a way to migrate a partition, all these activities require careful planning and highly skilled people, and often cause a significant downtime. In some cases, an SLA may be so strict that planned outages are not tolerated.

1.3 Live Partition Mobility is the answer

The Live Partition Mobility offered on POWER6 based systems is designed to enable the migration of an entire logical partition from one system to another. Live Partition Mobility uses a simple and automated procedure that transfers the configuration from source to destination without disrupting the hosted applications or the setup of the operating system and applications.

Live Partition Mobility provides the administrator a greater control over the usage of resources in the data center. It allows a level of reconfiguration that in the past was not possible due to complexity or just because of service level agreements that do not allow an application to be stopped for an architectural change.

The migration process can be performed either with a powered off or a live partition. There are two available migration types:

Inactive migration   The logical partition is powered off and moved to the destination system.

Active migration   The migration of the partition is performed while service is provided, without disrupting user activities.

1.3.1 Inactive migration

Inactive migration moves the definition of a powered off logical partition from one system to another along with its network and disk configuration. No additional change in network or disk setup is required and the partition can be activated as soon as migration is completed.

The inactive migration procedure takes care of the reconfiguration of involved systems.

- ► A new partition is created on the destination system with the same configuration present on the source system.
- ► Network access and disk data is preserved and made available to the new partition.
- ► On the source system, the partition configuration is removed and all involved resources are freed.

If a system is down due to scheduled maintenance or not in service for other reasons, an inactive migration may be performed. It is executed in a controlled way and with minimal administrator interaction so that it can be safely and reliably performed in a very short time frame.

When the service provided by the partition cannot be interrupted, its relocation can be performed with no loss of service by using the active migration feature.

1.3.2 Active migration

Using active migration, a running partition is moved from a source system to a destination system with no disruption of partition operation or user service.

An active migration performs the same operations as an inactive migration except that the operating system, the applications, and the services they provide are not stopped during the process. The physical memory content of the logical partition is copied from system to system allowing the transfer to be imperceptible to users.

During an active migration, the applications continue to handle their normal workload. Disk data transactions, running network connections, user contexts, and the complete environment are migrated without any loss and migration can be activated any time on any production partition.

There is no limitation on a partition's computing and memory configuration and multiple migrations can be executed concurrently. Both inactive and active migrations may involve partitions with any processing unit and memory size configuration.

1.4 Architecture

Live Partition Mobility requires a specific hardware infrastructure. Several platform components are involved.

1.4.1 Hardware infrastructure

The primary requirements for the migration of a logical partition are:

► Two POWER6 based systems controlled by the same Hardware Management Console (HMC).

    An optional redundant HMC configuration is supported.

► The destination system must have enough CPU and memory resources to host the mobile partition (the partition profile that is running, as alternate production profiles may exist).

► The operating system, applications, and data of the mobile partition must reside on virtual storage on an external storage subsystem.

► No physical adapters may be used by the mobile partition.

► The mobile partition's network and disk access must be virtualized using one or more Virtual I/O Servers.

    – The Virtual I/O Servers on both systems must have a shared Ethernet adapter configured to bridge to the same Ethernet network used by the mobile partition.

    – The Virtual I/O Servers on both systems must be capable of providing virtual access to all disk resources the mobile partition is using.

Live Partition Mobility requires a specific hardware and microcode configuration that is currently available on POWER6 based systems only.

The procedure that performs the migration identifies the resource configuration of the mobile partition on the source system and then reconfigures both source and destination systems accordingly. Since the focal-point of hardware configuration is the HMC, it has been enhanced to coordinate the process of migrating partitions. The same HMC must be used to manage both systems.

The mobile partition's configuration is not changed during the migration. The destination system must be able to host the mobile partition and must have enough free processor and memory resources to satisfy the partition's requirements before migration is started. There is no limitation on the size of the mobile partition; it can even use all resources of the source system offered by the Virtual I/O Server.

The operating system and application data must reside on external disks of the source system since the mobile partition's disk data must be available after the migration to the destination system is completed. An external, shared access storage subsystem is therefore required.

The mobile partition must not own any physical adapters and must use the Virtual I/O Server for both network and external disk access. The migration of partitions using multiple Virtual I/O Servers is supported.

Since the mobile partition's external disk space must be available to the Virtual I/O Servers on the source and destination systems, you cannot use storage pools. Each Virtual I/O Server must create virtual target devices using physical disks and not logical volumes.

Virtual network connectivity must be established before activating the partition migration task, while virtual disk setup is performed by the migration process.

Both the source and the target system must have an appropriate shared Ethernet adapter environment to host a moving partition. All virtual networks in use by the mobile partition on the source system must be available as virtual networks on the destination system.

VLANs defined by PVIDs on the VIOS have no meaning outside of an individual server as all packets are bridged untagged. It is possible for VLAN 1 on CEC 1 to be part of the 192.168.1 network while VLAN 1 on CEC 2 is part of the 10.1.1 network.

Because two networks are possible, you cannot look to see if VLAN 1 exists on both servers. You have to check if VLAN 1 maps to the same network on both servers.

Figure 1-1 shows a basic hardware infrastructure enabled for Live Partition Mobility. Each POWER6 based system is configured with a Virtual I/O Server partition. The mobile partition has only virtual access to network and disk resources. The Virtual I/O Server on the destination system is connected to the same network and is configured to access the same disk space used by the mobile partition. For illustration purposes, the device numbers are all shown as zero, but in practice, they will vary considerably.
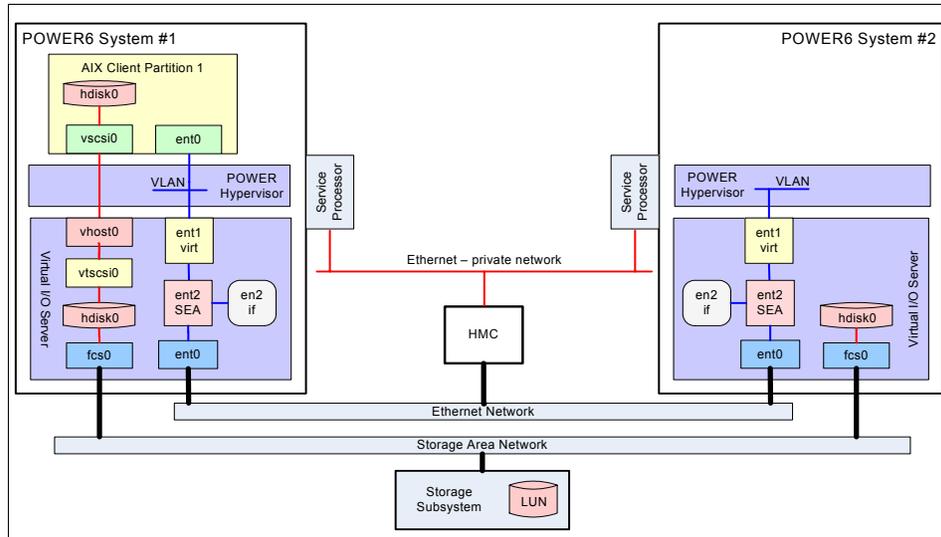
Figure 1-1   Hardware infrastructure enabled for Live Partition Mobility

The migration process creates a new logical partition on the destination system. This new partition uses the destination's Virtual I/O Server to access the same mobile partition's network and disks. During active migration, the state of the mobile partition is copied, as shown in Figure 1-2.
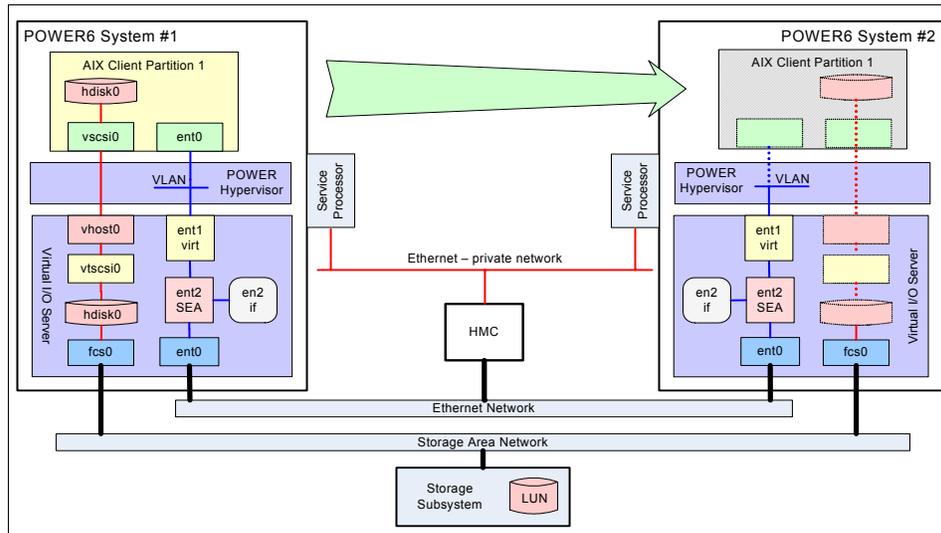
Figure 1-2   A mobile partition during migration

When the migration is complete, the source Virtual I/O Server is no longer configured to provide access to the external disk data. The destination Virtual I/O Server is set up to allow the mobile partition to use the storage. The final configuration is shown in Figure 1-3.
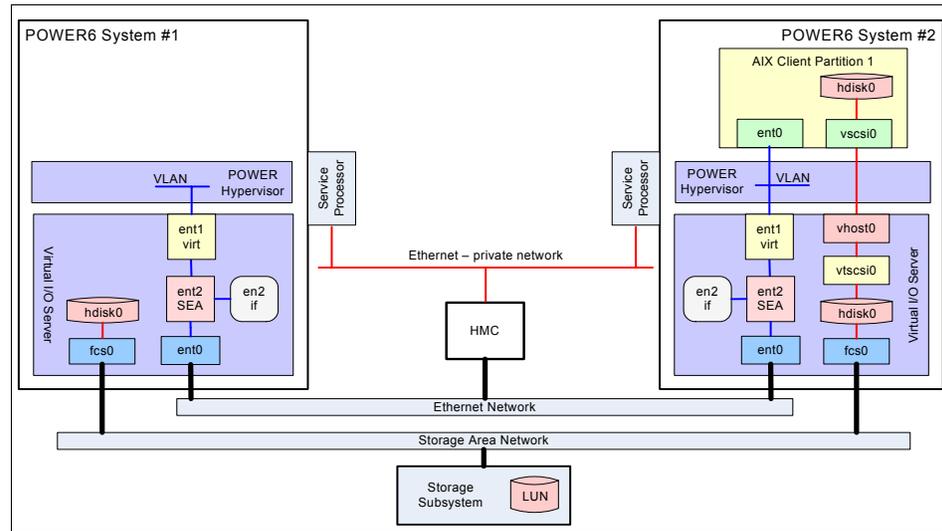
Figure 1-3   The final configuration after a migration is complete

1.4.2 Components involved

The Live Partition Mobility function changes the configuration of the two involved systems and, for active migration, manages the migration without interrupting the service provided by the applications running on the mobile partition.

The migration manager function resides on the HMC and is in charge of configuring both systems. It has the responsibility of checking that all hardware and software prerequisites are met. It executes the required commands on the two systems to complete migration while providing the migration status to the user.

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

Memory management of an active migration is assigned to a mover service partition on each system. During an active partition migration, the source mover service partition extracts the mobile partition's state from the source system and

it sends it over the network to the destination mover service partition, which in turn updates the memory state on the destination system.

Any Virtual I/O Server partition can be configured as a mover service partition.

Live Partition Mobility has no specific requirements on the mobile partition's memory size or the type of network connecting the mover service partitions. The memory transfer is a process that does not interrupt a mobile partition's activity and may take time when a large memory configuration is involved on a slow network. It is recommended to use high bandwidth connections, such as gigabit Ethernet.

1.5 Operation

Partition migration can be performed either as an inactive or an active operation.

1.5.1 Inactive migration

The basic steps required for inactive migration are:

1. Prepare the mobile partition for migration, if required, such as remove adapters that are not supported and ensure applications support mobility. See Appendix 2.2, "Live Partition Mobility prerequisites" on page 21 for additional information.

2. Shut down the mobile partition.

3. Perform the migration validation procedure provided by the HMC to verify that the migration can be performed successfully.

4. Start the inactive partition migration using the HMC.

   The HMC connects to both source and destination systems and performs the migration steps:

   a. It transfers the mobile partition's configuration from source to destination, including all partition profiles.

   b. It updates the destination Virtual I/O Server to provide virtual SCSI access to the mobile partition.

   c. It updates the source Virtual I/O Server to remove resources used to provide virtual SCSI access to the mobile partition.

   d. It removes the mobile partition configuration on the source system.

5. When migration is complete, the mobile partition can be activated on the destination system.

The steps executed are similar to those an administrator would follow when performing a manual migration. These actions normally require accurate planning and a system-wide knowledge of the configuration of the two systems since virtual adapters and virtual target devices have to be created on the destination system, following virtualization configuration rules.

The inactive migration task takes care of all planning and validation and performs the required activities without user action. This mitigates the risk of human error and executes the movement in a timely manner.

1.5.2 Active migration

The basic steps required for an active migration are:

1. Prepare the mobile partition for migration, keeping it active, such as remove adapters that are not supported and ensure applications support mobility. See Appendix 2.2, "Live Partition Mobility prerequisites" on page 21 for additional information..

2. Perform the migration validation procedure provided by the HMC to verify that the migration can be performed successfully.

3. Initiate the active partition migration using the HMC.

   The HMC connects to both source and destination systems and performs the migration steps:

   a. It transfers the mobile partition's configuration from source to destination, including all the partition's profiles.

   b. It updates the destination Virtual I/O Server to provide virtual SCSI access to the mobile partition.

   c. It activates the mover service partition function on the source and destination Virtual I/O Servers. The mover service partitions copy the mobile partition's state from the source to the destination system.

   d. It activates the mobile partition on the destination system.

   e. It updates the source Virtual I/O Server to remove resources used to provide virtual SCSI access to the mobile partition.

   f. It removes the mobile partition configuration on the source system.

Active migration performs similar steps to inactive migration, but also copies physical memory to the destination system. It keeps applications running, regardless of the size of the memory used by the partition; the service is not interrupted, the I/O continues to disk, and network connections keep transferring data.

1.6 Combining mobility with other features

There are many ways to take advantage of Live Partition Mobility. It can be exploited to perform actions that were not previously possible due to complexity or time constraints. Migration is a new function that can be combined with existing IBM System p features and software to provide better and more flexible service.

1.6.1 High availability clusters

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

For example, if a system has to be shut down due to a scheduled power outage, its hosted partitions may be migrated to powered systems before power is cut. This case is shown in Figure 1-4, where system A has to be shut down. The production database partition is actively migrated to system B, while the production Web application partition is actively migrated to system C. The test environment is not considered vital and is shut down during the outage.
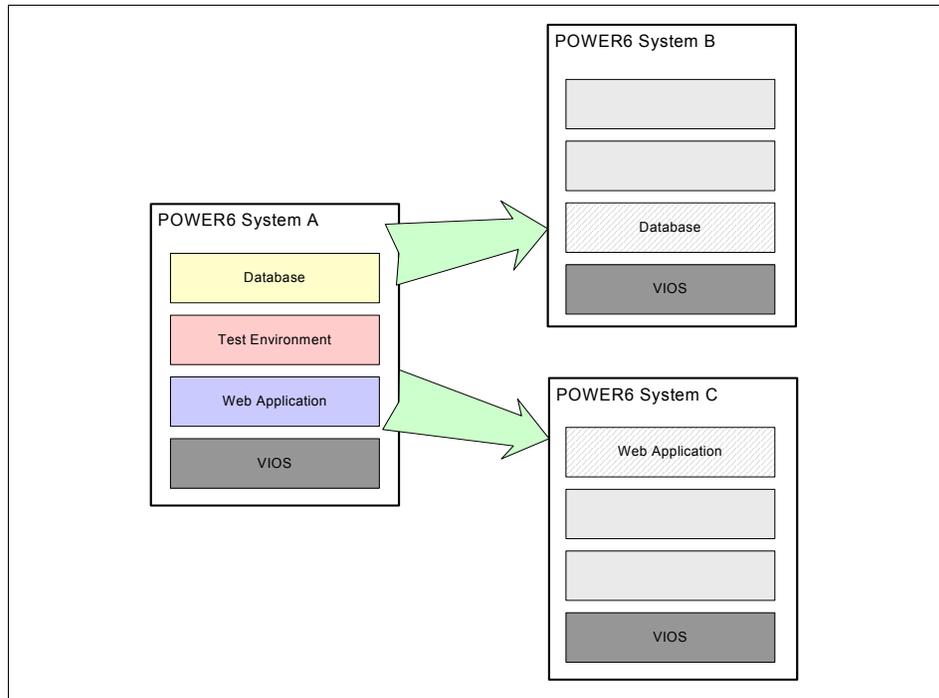
Figure 1-4   Migrating all partitions of a system

Live Partition Mobility is a reliable procedure for system reconfiguration and it may be used to improve the overall system availability.

High availability environments also require the definition of automated procedures that detect software and hardware events and activate recovery plans to restart a failed service as soon as possible.

Live Partition Mobility increases global availability, but it is not an high availability solution. It requires both source and destination systems be operational and that the partition is not in a failed state. In addition, it does not monitor operating system and application state and it is a user initiated action.

Unplanned outages still require specific actions that are normally executed by cluster solutions such as IBM HACMP.

Cluster software and Live Partition Mobility provide different functions that can be used together to improve the availability and uptime of applications. They can simplify administration, reducing the related cost.

1.6.2 AIX Live Workload Partitions

AIX Version 6 allows you to group applications running on the same AIX image, together with their disk data and network configuration. Each group is called a Workload Partition.

Live Workload Partitions are migration capable. Given two running AIX Version 6 images that share a common file system, the administrator can decide to actively migrate a workload between operating systems, keeping the applications running.

While workload partition migration is very similar to active Live Partition Mobility, it is a pure AIX function. It does not require any partition configuration change and it can be executed on any server running AIX Version 6, including POWER4™ and POWER5 based servers. Live Workload Partitions is a feature of AIX Version 6 and will function on all systems that support AIX Version 6, while Live Partition Mobility is an Advanced POWER Virtualization feature that works for all operating systems that operate on POWER6-based System p servers.

The Workload Partition migration function does not require a configuration of virtual devices in the source and destination systems. AIX keeps running on both systems and continues to use its allocated resources. It is the system administrator's task to perform a dynamic partition reconfiguration operation to reduce the footprint of the source partition and enlarge the destination partition. Workload Partition migration also requires the destination partition to exist and be running before it is started.

Figure 1-5 represents an example of Live Workload Partitions usage. System B is a POWER6 system with three different workloads. Each of them can be migrated to another AIX Version 6 image even if they run on different hardware platforms.
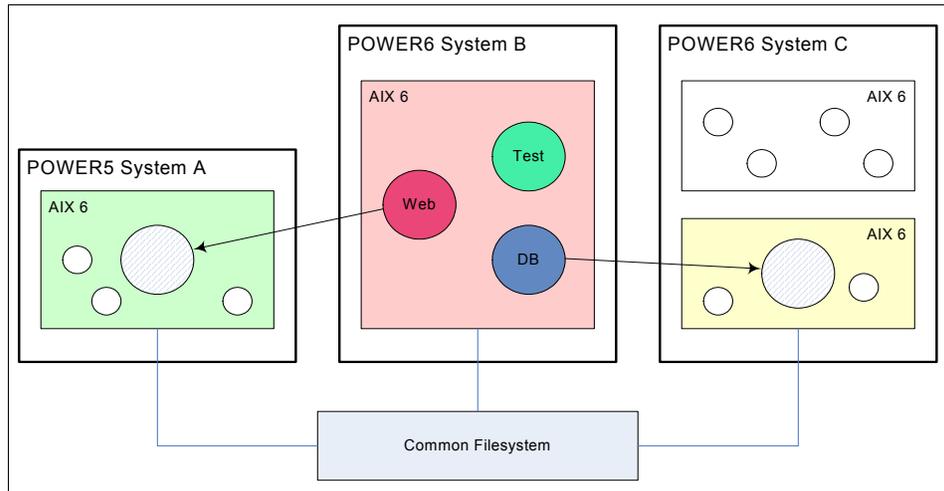
Figure 1-5   AIX Workload Partition example

Live Partition Mobility and AIX Workload Partition migration have different scopes while having similar characteristics. They can be used in conjunction to provide even higher flexibility in a POWER6 environment.

2

Live Partition Mobility mechanisms

Chapter 1, "Overview" on page 1 provides a high-level overview of Live Partition Mobility for inactive and active migrations.

This chapter presents the components involved in Live Partition Mobility and their respective roles. It discusses the compatibility, capability, and readiness of partitions and systems to participate in inactive and active migrations. It then goes on to describe the mechanisms of Live Partition Mobility in detail. The chapter concludes with some observations on the influence of the infrastructure of partition migration.

This discussion is provided within the following topics:

► 2.1, "Live Partition Mobility components" on page 18
► 2.2, "Live Partition Mobility prerequisites" on page 21
► 2.3, "Partition migration high-level workflow" on page 24
► 2.4, "Inactive partition migration" on page 25
► 2.5, "Active partition migration" on page 29
► 2.6, "Performance considerations" on page 38
► 2.7, "AIX and active migration" on page 39

17

2.1 Live Partition Mobility components

Inactive and active partition migration from one physical system to another is achieved through interaction between several components. These are shown in Figure 2-1.
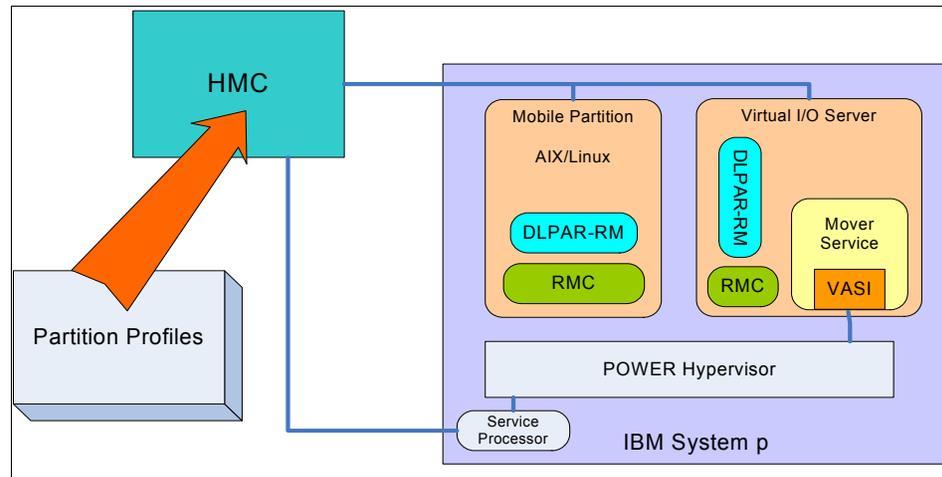
Figure 2-1   Live Partition Mobility components

These components and their roles are described as follows.

Hardware Management Console (HMC)

> The HMC is the central point of control. It coordinates administrator initiation and setup of the subsequent migration command sequences that flow between the various partition migration components.

> The HMC provides both a GUI wizard interface and a command-line interface to control migration. The HMC interacts with the service processors and POWER Hypervisor™ on the source and destination servers, the mover service partitions, the Virtual I/O Server partitions, and the mobile partition itself.

Resource Monitoring and Control (RMC)

> A distributed framework and architecture that allows the HMC to communicate with a managed logical partition.

Dynamic LPAR Resource Manager

> An RMC daemon that runs inside the AIX, Linux, and

Virtual I/O Server partitions. The HMC uses this capability to remotely execute partition specific commands.

Virtual asynchronous services interface (VASI)

The source and destination mover service partitions use this new virtual device type to communicate with the POWER Hypervisor to gain access to partition state.

The VASI device must be configured on both the source and the destination Virtual I/O Servers for the mobile partition to participate in an active partition migration. Since the VASI device is only used by the mover service partition, it is only required for active partition migration.

Mover service partition (MSP)

Mover Service Partition (MSP) is a new *attribute* of the Virtual I/O Server partition. It is a Virtual I/O Server partition that includes a function that asynchronously extracts, transports, and installs partition state. Two mover service partitions are involved in an active partition migration: one on the source system, the other on the destination system. Mover service partitions are not used for inactive migrations.

A mover service partition must contain a single VASI device. This may be added or removed from a partition's profile or moved in or out of a running Virtual I/O Server partition using dynamic reconfiguration.

POWER Hypervisor Active partition migration requires server hypervisor support to process both informational and action requests from the HMC and to transfer a partition state through the VASI in the mover service partitions.

Virtual I/O Server (VIOS)

Only virtual adapters can be migrated with a partition. The physical resources that back the mobile partition's virtual adapters must be accessible by the Virtual I/O Servers on both the source and destination systems.

Partition profiles The HMC copies all of the mobile partition's profiles without modification to the target system as part of the migration process.

The HMC creates a new migration profile containing the partition's current state. Unless you specify a profile name when the migration is started, this profile replaces the existing profile that was last used to activate the partition. If you specify an existing profile name, the HMC replaces

that profile with the new migration profile. Therefore, if you do not want the migration profile to replace any of the partition's existing profiles, you must specify a new, unique profile name when starting the migration. All profiles belonging to the mobile partition are deleted from the source server once the migration has completed.

If the mobile partition's profile is part of a system profile on the source server, then it is automatically removed once the source partition is deleted. It is not automatically added to a system profile on the target server.

Time reference   Time reference is a new attribute of partitions, including Virtual I/O Server partitions. This partition attribute is only supported on managed systems that are capable of active partition migration.

Synchronizing the time of day clocks for the source and destination Virtual I/O Server partitions is optional for both active and inactive partition migration. However, it is a recommended step for active partition migration. If you choose not to complete this step, the source and destination systems will synchronize the clocks while the mobile partition is moving from the source system to the destination system.

The time reference partition (TRP) setting has been introduced to enable the POWER Hypervisor to synchronize the mobile partition's time of day as it moves from one system to another. It uses Coordinate Universal Time (UTC) derived from a common network time protocol (NTP) server with NTP clients on the source and destination systems. More than one TRP can be specified per system. The POWER Hypervisor uses the longest running time reference partition as the provider of authoritative system time. It can be set or reset through POWER Hypervisor while the partition is running.

2.1.1 Other components affecting Live Partition Mobility

Though not considered to be part of Live Partition Mobility, some other System p server components may influence, or are influenced by, the mobility of a partition.

Integrated Virtual Ethernet adapter (IVE)
A host Ethernet adapter uses a two or four-port integrated Ethernet adapter, directly attached to the POWER Hypervisor. The hypervisor can create up to 32 logical

Ethernet ports that can be given to one or more logical partitions. This provides a partition with a virtual Ethernet communications link, seen as a logical host Ethernet adapter (LHEA), without recourse to a shared Ethernet adapter in a Virtual I/O Server.

Performance monitor API (PMAPI)
An AIX subsystem comprising commands, libraries, and a kernel extension that controls the use of the POWER performance registers.

Barrier Synchronization Registers (BSR)
Barrier synchronization registers provide a fast, lightweight barrier synchronization between CPUs.

This facility is intended for use by application programs that are structured in a single instruction, multiple data (SIMD) manner. Such programs often proceed in phases where all tasks synchronize processing at the end of each phase. The BSR is designed to accomplish this efficiently. Barrier synchronization registers are not migratable and cannot be reconfigured dynamically.

2.2 Live Partition Mobility prerequisites

Live Partition Mobility requires coordinated movement of a partition's state and resources. Migratable partitions move between capable, compatible, and ready systems.

A single HMC can control several concurrent migrations. There are no architectural restrictions on the number of migrations that can be underway at any one time. However, a single mover service partition can handle a maximum of four simultaneous active migrations. It is possible to have several mover service partitions on a system. In practice, the maximum number of concurrent migrations is limited by the processing capacity of the HMC and contention for HMC locks.

2.2.1 Capability and compatibility

The first step of any mobility operation is to validate the capability and compatibility of the source and destination systems. The high-level prerequisites for Live Partition Mobility are:

► A migration-capable ready source system.

- ► A migration-capable ready destination system.

- ► Compatibility between the source and destination systems.

- ► The source and destination systems must be under the control of the same HMC.

- ► A migratable, ready partition to be moved from the source system to the destination system. For an inactive migration, the partition must be powered down, but must be capable of booting on the destination system.

- ► For active migrations, a mover service partition on the source and destination systems.

- ► One or more storage area networks (SAN) that provide connectivity to all of the mobile partition's disks to the Virtual I/O Server partitions on both the source and destination servers. The mobile partition accesses all migratable disks through virtual SCSI devices. The LUNs must be zoned and masked to the Virtual I/O Servers on both systems. SCSI reservation must be disabled.

- ► The mobile partition's virtual disks must be mapped to LUNs; they cannot be part of a storage pool or logical volume on the Virtual I/O Server.

- ► One or more physical IP networks (LAN) that provide the necessary network connectivity for the mobile partition through the Virtual I/O Server partitions on both the source and destination servers. The mobile partition accesses all migratable network interfaces through virtual Ethernet devices.

- ► An RMC connection to manage inter-system communication.

If any of the above elements are missing, a migration cannot occur.

Before initiating the migration of a partition, the HMC verifies the capability and compatibility of the source and destination servers along with the characteristics of the mobile partition to determine whether or not a migration is possible.

The hardware, firmware, Virtual I/O Server, mover service partition, operating system, and HMC versions that are required for Live Partition Mobility along with the system compatibility requirements are described in Chapter 3, "Requirements and preparation" on page 41.

2.2.2 Readiness

Migration readiness is a dynamic partition property that changes over time.

Server readiness

A server that is running on battery power is not ready to receive a mobile partition; it cannot be selected as a destination for partition migration. A server

that is running on battery power may be the source of a mobile partition; indeed, that it is running on battery power may be the impetus for starting the migration.

Infrastructure readiness

A migration operation requires a SAN and a LAN to be configured with their corresponding virtual SCSI, VLAN, and virtual Ethernet devices. At least one Virtual I/O Server on both the source and destination systems must be configured as a mover service partition with a VASI device for active migrations. The HMC must have RMC connections to the Virtual I/O Servers and a connection to the service processors on the source and destination servers. For an active migration, the HMC also needs RMC connections to the mobile partition and the mover service partitions.

2.2.3 Migratability

The term *migratability* refers to a partition's ability to be migrated and is distinct from partition readiness. A partition may be migratable but not ready. A partition that is not migratable may be made migratable with a configuration change. For active migration, consider whether a shutdown and reboot is required. When considering a migration, you should bear in mind the following, in addition to the prerequisites.

- ► General prerequisites:
  - The memory and processor resources required to meet the mobile partition's current entitlements must be available on the destination server.
  - The partition must not have any required dedicated physical adapters.
  - The partition is not a Virtual I/O Server.
  - The partition is not designated as an redundant error path reporting partition.
  - The partition does not have any of its virtual SCSI disks defined as logical volumes in any Virtual I/O Server. All virtual SCSI disks must be mapped to LUNs visible on a SAN.
  - The partition is not part of an LPAR workload group. A partition can be dynamically removed from a group.
  - The partition has a unique name. A partition cannot be migrated if any partition exists with the same name on the destination server.
- ► The following characteristics apply to inactive migration only:
  - A partition in the Not Activated state.
  - May use huge pages
  - May use the barrier synchronization registers.

► The following characteristics apply to active migration only:

– A partition that is in the Running state.

– Active partition migration does not support non-default virtual serial I/O adapters, unless the virtual serial I/O is marked as desired, and can be dynamically removed from the partition. The removal of any virtual serial I/O must be done prior to starting a migration.

The two default server serial adapters that are automatically created and assigned to a partition when a partition is created are automatically recreated on the destination system by the migration process.

2.3 Partition migration high-level workflow

The steps of inactive and active partition migration follow the same four-step sequence:

1. Preparation

   Ready the infrastructure to support Live Partition Mobility.

2. Validation

   Check the configuration and readiness of the source and destination systems.

3. Migration

   Transfer of partition state from the source to destination takes place. The same command is used to launch inactive and active migrations. The HMC determines the appropriate type of migration to use based on the state of the mobile partition.

   – If the partition is in the Running state, then the migration is active.

   – If partition is in the Not Activated state, then the migration is inactive.

4. Completion

   Free unused resources on the source system and the HMC.

The remainder of this chapter describes the inactive and active migration processes.

Note: As part of a migration process, the HMC copies all of the mobile partition's profiles as-is to the destination system. The HMC also creates a new migration profile containing the partition's current state and, unless you specify a profile name, this profile replaces the existing profile that was last used to activate the partition. If you specify an existing profile name, the HMC replaces that profile with the new migration profile. Therefore, if you want to keep the partition's existing profiles, you should specify a new and unique profile name when initiating the migration. If you add an adapter (physical or virtual) to a partition using dynamic reconfiguration, it is added to the profile as desired.

2.4 Inactive partition migration

Inactive partition migration allows you to move a powered off partition, along with its profiles and virtualized resources, from one server to another. The mobile partition retains its name, its inactive state, and its NVRAM. Its virtual I/O resources are assigned and remapped to the appropriate Virtual I/O Server partitions on the destination system. Its processor and memory resources remain unassigned until it is activated.

2.4.1 Introduction

The HMC is the central point of control, coordinating administrator actions and migration command sequences. Because the mobile partition is powered off, only the static partition state (definitions and configurations) is transferred from source to destination. The transfer is performed by the controlling HMC, the service processors, and the POWER hypervisor on the two systems; there is no dynamic state, so mover service partitions are not required.

The HMC creates a migration profile for the mobile partition on the destination server corresponding to its current configuration. All the profiles associated with the mobile partition are moved to the destination after the partition definition has been created on the destination.

Note: Because the HMC always migrates the latest activated profile, an inactive partition that has never been activated is not migratable. To meet this requirement, it is not necessary to boot to an operating system; booting to the SMS menu is sufficient. Any changes to the latest activated profile after power off will not be preserved. In order to keep these changes, the mobile partition must be reactivated and shut down.

2.4.2 Validation phase

The HMC performs a pre-check that ensures that you are performing a valid migration, that there are no high-level blocking problems, and that the migration has a good chance of being successful. The validation workflow is schematically shown in Figure 2-2.
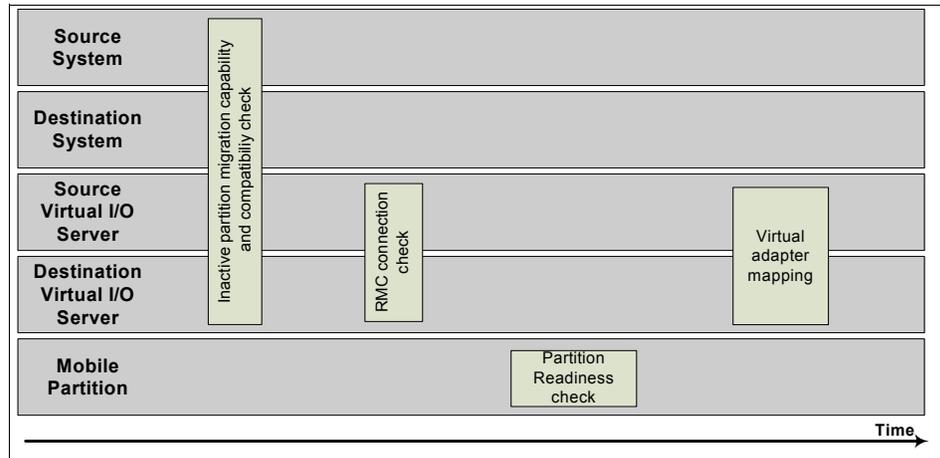
Figure 2-2   Inactive migration validation workflow

The inactive migration validation process performs the following operations:

► Checks the Virtual I/O Server and hypervisor migration capability and compatibility on the source and destination.

► Checks that resources (processors, memory, and virtual slots) are available to create a shell partition on the destination system with the exact configuration of the mobile partition.

► Verifies the RMC connections to the source and destination Virtual I/O Servers.

► Ensures that the partition name is not already in use at the destination.

► Checks for virtual MAC address uniqueness.

► Checks that the partition is in the Not Activated state.

► Ensures that the mobile partition is an AIX or Linux partition, is not an alternate path error logging partition, is not a service partition, and is not a member of a workload group.

► Ensures that the mobile partition has an active profile.

► Checks the number of current inactive migrations against the number of supported inactive migrations.

- ► Checks that all required I/O devices are connected to the mobile partition through a Virtual I/O Server, that is, there are no required physical adapters.
- ► Verifies that the virtual SCSI disks assigned to the partition are accessible by the Virtual I/O Servers on the destination system.
- ► Creates the virtual adapter migration map that associates adapters on the source Virtual I/O Servers with adapters on the destination Virtual I/O Servers.
- ► Ensures that no virtual SCSI disks are backed by logical volumes and that no virtual SCSI disks are attached to internal disks (not on the SAN).

2.4.3 Migration phase

If all the pre-migration checks pass, the migration phase can start. For inactive partition migration, the transfer of state follows a path:

1. From the source hypervisor to the HMC
2. From the HMC to the destination hypervisor

This is shown in Figure 2-3.
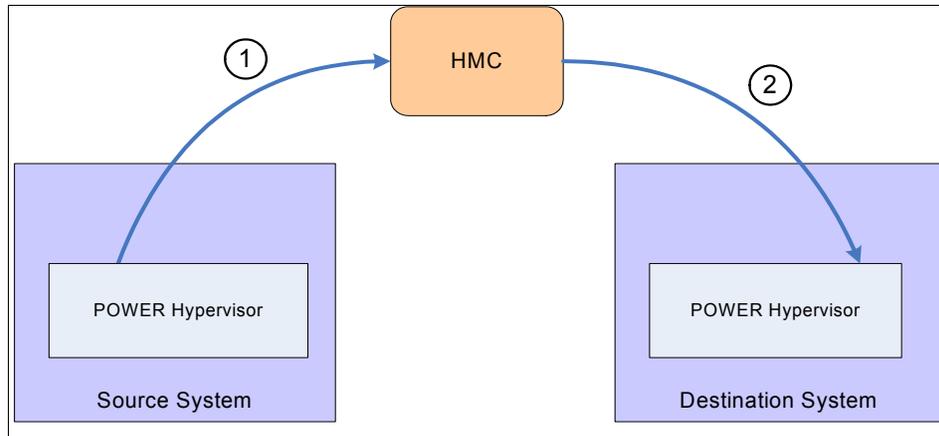
Figure 2-3   Inactive migration state flow

The inactive migration workflow is shown in Figure 2-4.
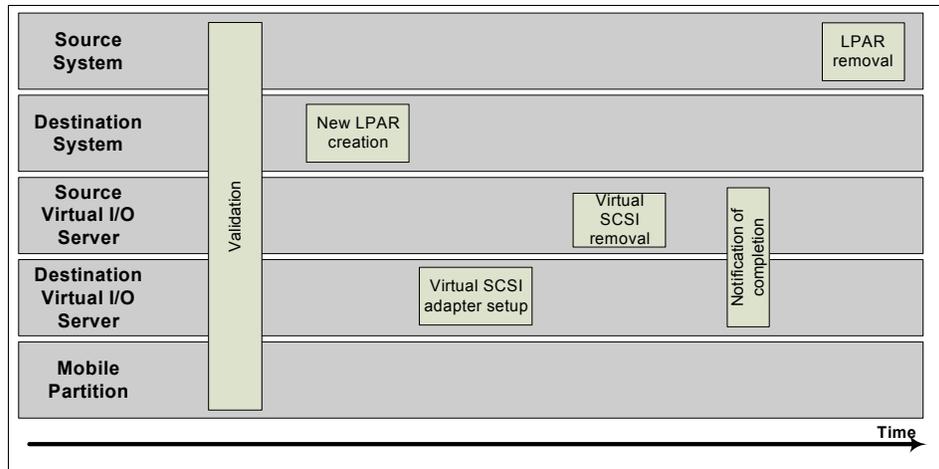
Figure 2-4   Inactive migration workflow

The details of this workflow are provided as follows:

1. The HMC inhibits any changes to the source system and the mobile partition that might invalidate the migration.

2. The HMC extracts the virtual device mappings from the source Virtual I/O Servers and uses this to generate a source-to-destination virtual adapter migration map. This map ensures that there is no loss of multipath I/O capability for virtual SCSI and virtual Ethernet. The HMC fails the migration request if the device migration map is incomplete.

3. The HMC creates a compatible partition shell on the destination system.

4. The HMC creates a migration profile for the mobile partition's current (last-activated) profile. If the mobile partition was last activated with profile *my_profile* and resources were moved in to or out of the partition before the partition was shutdown, then the migration profile will differ from that of *my_profile*.

5. The HMC copies over the partition profiles. This includes all the existing profiles associated with the mobile partition on the source system and the migration profile. The existing partition profiles are not modified at all during the migration - the virtual devices are not re-mapped to the new system.

6. The HMC creates the required virtual SCSI adapters in the Virtual I/O Servers on the destination system and completes the LUN to virtual adapter mapping.

7. On completion of the transfer of state, the HMC sets the migration state to completed and informs the POWER hypervisor on both the source and destination.

2.4.4 Migration completion phase

When the migration is complete, unused resources are deleted.

1. The source Virtual I/O Servers remove the virtual adapter slots and virtual target devices used by the mobile partition. The HMC removes the virtual slots from the source Virtual I/O Server's profile.

2. The HMC deletes the partition on the source server.

> **Note:** Virtual slot numbers can change during migration. When moving a partition to a server and then back to the original, it will not have the same slot numbers. If this information is required, you should record slot numbers.

2.4.5 Stopping an inactive partition migration

You can stop an inactive partition migration from the controlling HMC while the partition is in the Migration starting state. The HMC performs automatic rollback of all reversible changes and identifies all non-reversible changes.

2.5 Active partition migration

The active partition migration function provides the capability to move a running operating system, hosted middleware, and applications between two systems without disrupting the service provided. Databases, application servers, network and SAN connections, and user applications are all transferred in a manner transparent to users. The mobile partition retains its name, its active state, its NVRAM, its profiles, and its current configuration. Its virtual I/O resources are assigned and remapped to the appropriate Virtual I/O Server partitions on the destination system.

2.5.1 Active partition state

In addition to the partition definition and resource configuration, active migration involves the transfer of active run-time state. This state includes:

► The partition's memory

► The hardware page table (HPT)

► The processor state

► The virtual adapter state

► The non-volatile RAM (NVRAM)

- ► The time of day (ToD)
- ► The partition configuration
- ► The state of each resource

It is the mover service partitions on the source and destination, under the control of the HMC, that move this state between the two systems.

2.5.2 Preparation

Once you have created the Virtual I/O Servers, mover service partitions, and VASI devices, you must prepare the source and destination systems for migration.

1. Synchronize the time of day clocks on the mover service partitions using an external time reference, such as the network time protocol (NTP). This is an optional step that increases the accuracy of time measurement during migration. It is not required by the migration mechanisms and even if this step is omitted, the migration process correctly adjusts the partition time. Time never goes backwards on the mobile partition during a migration.

2. Prepare the partition for migration.

   – Use dynamic reconfiguration on the HMC to remove all dedicated I/O, such as PCI slots, GX slots, and HEA from the mobile partition.

   – Remove the partition from a partition workload group.

3. Prepare the destination Virtual I/O Server.

   – Configure the shared Ethernet adapter as necessary to bridge VLANs.

   – Configure the SAN such that requisite storage devices are available.

4. Initiate the partition migration using either the GUI, or CLI on the HMC.

   Select:

   – The partition to migrate.

   – The destination system.

   – Optionally, the mover service partitions on the source and destination systems.

     If there is only one active mover service partition on the source or the destination server, then the mover service partition selection is automatic. If there are multiple active mover service partitions on one or both, then you can either specify which ones to use, or let the HMC choose for you.

   – Optionally, the virtual device mappings in the destination Virtual I/O Server. See 5.6, "The command-line interface (CLI)" on page 138 for the details.

2.5.3 Validation phase

Once the source and destination mover service partitions have been identified, the HMC performs a pre-check to ensure that the user is performing a valid migration, that there are no high-level blocking problems, and that the environment satisfies the prerequisites for a migration operation.

After the pre-check, the HMC prevents any configuration changes to the partition that might invalidate the migration and then proceeds to perform a detailed capability, compatibility, migratability, and readiness check on the source and destination systems.

All configuration checks are performed during each validation to provide a complete list of potential problems.

The workflow for the active migration validation is shown in Figure 2-5.
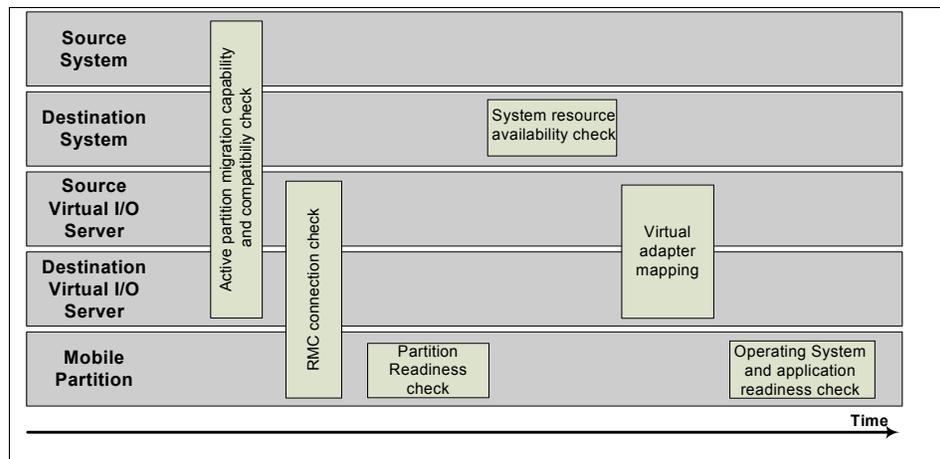
Figure 2-5   Active migration validation workflow

Configuration checks
The HMC performs the following configuration checks.

► Checks the source and destination systems, POWER hypervisor, Virtual I/O Servers, and mover service partitions for active partition migration capability and compatibility.

► Checks that the RMC connections to the mobile partition, the source and destination Virtual I/O Servers, and the connection between the source and destination mover service partitions are established.

► Checks that there are no required physical adapters in the mobile partition and that there are no required virtual serial slots higher than slot 2.

- ► Checks that no client virtual SCSI disks on the mobile partition are backed by logical volumes and that no disks map to internal disks.

- ► Checks the mobile partition, its OS, and its applications for active migration capability. An application registers is capability with AIX, and may block migrations.

- ► Checks that the logical memory block size is the same on the source and destination systems.

- ► Ensures that the type of the mobile partition is aixlinux and that it is neither an alternate error logging partition nor a mover service partition.

- ► Ensures that the mobile partition is not configured with barrier synchronization registers.

- ► Ensures that the mobile partition is not configured with huge pages.

- ► Checks that the partition state is active/running.

- ► Checks that the mobile partition is not in a partition workload group.

- ► Checks the uniqueness of the mobile partition's virtual MAC addresses.

- ► Checks that the mobile partition's name is not already in use on the destination server.

- ► Checks the number of current active migrations against the number of supported active migrations.

Resource availability checks

After verifying system and partition configurations, the HMC determines whether or not sufficient resources are available on the destination server to host the inbound mobile partition. The following are performed:

1. A check that the necessary resources (processors, memory, and virtual slots) are available to create a shell partition on the destination system with the exact configuration of the mobile partition.

2. The HMC generates a source-to-destination hosting virtual adapter migration map, ensuring no loss of multipath I/O capability for virtual SCSI and virtual Ethernet. The HMC fails the migration request if the device migration map is incomplete.

3. The HMC instructs the operating system in the mobile partition to check its own capacity and readiness for migration. AIX passes the check-migrate request to those applications and kernel extensions that have registered to be notified of dynamic reconfiguration events. The operating system either accepts or rejects the migration. In the latter case, the HMC fails the migration request.

This is the end of the validation phase. At this point, there have been no state changes to the source and destination systems or to the mobile partition. The HMC inhibits all further dynamic reconfiguration of the mobile partition that might invalidate the migration: CPU, memory, slot, variable capacity weight, processor entitlement, and LPAR group.

The partition migration phase is ready to start.

2.5.4 Partition migration phase

If all the validation checks pass, then the HMC initiates the migration procedure. From this point forward, all state changes are rolled back in the event of an error.

Figure 2-6 shows the activities and workflow of the migration phase of an active migration.
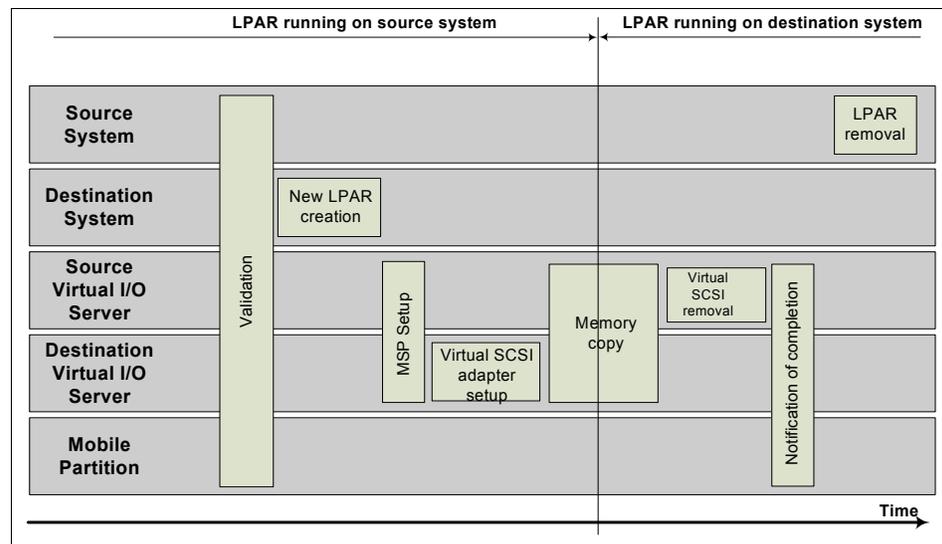
Figure 2-6   Migration phase of an active migration

For active partition migration, the transfer of partition state follows a path:

1. From the mobile partition to the source system's hypervisor.

2. From the source system's hypervisor to the source mover service partition.

3. From the source mover service partition to the destination mover service partition.

4. From the destination mover service partition to the destination system's hypervisor.

5. From the destination system's hypervisor to the partition shell on the destination.

This is shown graphically in Figure 2-7.
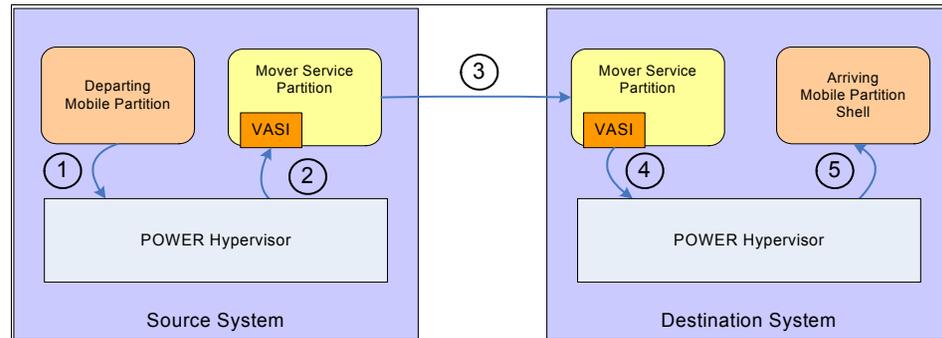
Figure 2-7   Active migration partition state transfer path

The migration process comprises the following steps:

1. The HMC creates a compatible partition shell on the destination system. This shell partition is used to reserve the resources required to receive the inbound mobile partition.

   The pending values of the mobile partition are not preserved across the migration; the current values of the partition on the source system become both the pending and the current values of the partition on the destination system. The configuration of the partition on the source system includes:

   – Processor configuration: dedicated or shared processors, processor counts, and entitlements (minimum, maximum, and desired)

   – Memory configuration (minimum, maximum, and desired)

   – Virtual adapter configuration

   The creation of the partition shell on the destination system ensures that all of the required resources are available for the mobile partition and cannot be stolen at some point during the migration. The current partition profile associated with the mobile partition is created on the destination system.

2. The HMC configures the mover service partitions on the source and destination systems. These two movers establish:

   – A connection to their respective POWER hypervisor through the VASI adapter

   – A private, full-duplex communications channel between themselves, over a standard TCP/IP connection, for transporting the moving partition's state

3. The HMC issues a prepare for migration event to the migrating operating system (still on the source system), giving the mobile partition the opportunity to get ready to be moved. The operating system passes this event to registered kernel extensions and applications, so that they may take any necessary actions, such as reducing memory footprint, throttling workloads, loosening heartbeats, and other timeout thresholds. The operating system inhibits access to the PMAPI registers and zeroes internal counters upon receipt of this event.

   If the partition is not ready to perform a migration at this time, then it returns a failure indicator to the HMC, which cancels the migration and rolls back all changes.

4. The HMC creates the virtual target devices and virtual SCSI adapters in each of the Virtual I/O Servers on the destination system that will host the virtual SCSI client adapters of the mobile partition. This step uses the virtual adapter migration map created during the validation phase. The migration is stopped in the event of an error.

5. The mover on the source system starts sending the partition state to the mover on the destination system, copying the mobile partition's physical pages to the physical memory reserved by the partition shell on the destination.

6. Since the mobile partition is still active, with running applications, its state continues to change while the memory is being moved from one system to the other. Memory pages that are modified during the transfer of state are marked modified, or *dirty*. After the first pass, the source mover re-sends all the dirty pages. This process is repeated until the number of pages marked as dirty at the end of each loop no longer decreases, or is considered sufficiently small, or a timeout is reached.

   Based on the total number of pages associated with partition state and the number of pages left to transmit, the mover service partition instructs the hypervisor on the source system to suspend the mobile partition.

7. The mobile partition confirms the suspension by quiescing all its running threads.

   The partition is now suspended.

8. During the partition suspension, the source mover service partition continues to send partition states to the destination server.

   The partition is now resumed.

9. The mobile partition resumes execution on the destination server, re-establishing its operating environment. This is the point of no return; the migration can no longer be rolled back to the source. If the migration fails after this, recovery will complete the migration on to the destination system.

It is possible that the mobile partition resumes execution before all its memory pages have been copied to the destination. If the mobile partition requires a page that has not yet been migrated, then it is demand-paged from the source system. This technique significantly reduces the length of the pause, during which the partition is unavailable.

10. The mobile partition recovers I/O, retrying all pending I/O requests that were not completed while on the source system. It also sends a gratuitous ARP requests on all VLAN virtual adapters to update the ARP caches in the various switches and systems in the external network.

   The partition is now active and visible again.

11. When the destination mover service partition receives the last dirty page from the source system, the migration is complete.

The period between the end of step 7 and the end of step 10 in the preceding list is called the *suspend window*. It lasts just a few seconds.

2.5.5 Migration completion phase

The final steps of the migration return all of the resources to the source and destination systems and restore the partition to its fully functional state.

1. With the completion of the state transfer, the communications channel between the two mover service partitions is closed along with their VASI connections to their respective POWER hypervisor.

2. The Virtual I/O Servers on the source system remove the virtual SCSI server adapters associated with the mobile partition by:

   – Unlocking the virtual SCSI server adapters

   – Removing the device-to-LUN mappings

   – Closing the device drivers

   – Deleting the virtual SCSI server adapters

3. On the mobile partition, AIX notifies all registered kernel extensions and applications that the migration is complete so that they may perform any required recovery operations.

4. The HMC informs the source and destination mover service partitions that the migration is complete and that they can delete the migration data from their tables.

5. The HMC deletes the mobile partition and all its profiles on the source server.

6. You can now add dedicated I/O adapters as required using dynamic reconfiguration and add the mobile partition to an LPAR group.

2.5.6 Virtual I/O Server selection

If the HMC cannot find a virtual adapter mapping for a migration, then the migration is halted at the validation phase.

The HMC must identify at least one possible destination Virtual I/O Server for each virtual SCSI client adapter assigned to the mobile partition, or it will fail the pre-check or migration. The destination Virtual I/O Servers must have access to all the LUNs used by the mobile partition. If multiple source to destination Virtual I/O Server combinations are possible for virtual adapter mappings, and you have not specified a mapping, the HMC selects one of them.

Suggested mappings are given if the following criteria are met:

► The mobile partition's virtual client SCSI adapters that are assigned to a single Virtual I/O Server on the source server will be assigned to a single Virtual I/O Server on the destination system.

► The mobile partition's virtual SCSI client adapters that are assigned to two or more different Virtual I/O Servers on the source system will be assigned to the same number of Virtual I/O Servers on the destination system.

Failure to find a suggested good match during the separately-run migration pre-check happens if either of the following is true:

► The mobile partition's virtual SCSI client adapters that are currently assigned to a single Virtual I/O Server on the source server will have to be assigned to different Virtual I/O Servers on the destination system.

► The mobile partition's virtual SCSI client adapters that are currently assigned to different Virtual I/O Server on the source server will have to be assigned to a single Virtual I/O Server on the destination.

If the destination Virtual I/O Servers cannot access all the VLANs required by the mobile partition, the HMC halts the migration.

Both possible and suggested HMC-selected Virtual I/O Servers, if they exist, are viewable on the HMC with through the graphical user interface and the HMC `lslparmigr -r virtualio` command.

2.5.7 Source and destination mover service partition selection

The HMC selects a source and destination mover service partition unless you provide them explicitly. If no movers are available, the migration fails. Valid source and destination mover service partition pairs that can be used for a migration can be seen on the HMC with the graphical user interface and the HMC `lslparmigr -r msp` command.

If either of the chosen mover service partitions determines that its VASI cannot handle a migration or if the HMC receives a VASI device error from a mover service partition, the HMC stops the migration with an error.

2.5.8 Stopping an active migration

You can stop an active partition migration through the controlling HMC while the mobile partition is in the Migration starting state. If stopped during the allowable window, the partition remains on the source system as though the migration had not been started. The allowable window is while the partition is in the Migration starting state on the source system. If you try to stop a migration after the Migration starting state, then the HMC takes no action other than displaying an error message.

If the source or destination server is powered down after the HMC has enabled suspension on the mobile partitions, the HMC must stop the migration and perform a rollback of all reversible changes. When the hypervisor resumes operation, the partitions come back up in the powered off state, with a migration state of invalid.

2.6 Performance considerations

Active partition migration involves moving the state of a partition from one system to another while the partition is still running. The mover service partitions working with the hypervisor use partition virtual memory functions to track changes to partition memory state on the source system while it is transferring memory state to the destination system.

During the migration phase, there is an initial transfer of the mobile partition's physical memory from the source to the destination. Since the mobile partition is still active, a portion of the partition's resident memory will almost certainly have changed during this pass. The hypervisor keeps track of these changed pages for retransmission to the destination system in a dirty page list. It makes additional passes through the changed pages until the mover service partition detects that a sufficient amount of pages are clean or the timeout is reached.

The speed and load of the network used to transfer state between the source and destination systems influence the time required for both the transfer of the partition state and the performance of any remote paging operations.

The amount of changed resident memory after the first pass is controlled more by write activity of the hosted applications than by the total partition memory size. Nevertheless, it is reasonable to assume that partitions with a large memory

requirement will have higher numbers of changed resident pages than smaller ones.

To ensure that active partition migrations are truly non-disruptive, even for large partitions, the POWER Hypervisor resumes the partition on the destination system before all the dirty pages have been migrated over to the destination. If the mobile partition tries to access a dirty page that has not yet been migrated from the source system, the hypervisor on the destination sends a demand paging request to the hypervisor on the source to fetch the required page.

Providing a high-performance network between the source and destination mover partitions and reducing the partition's memory update activity prior to migration will improve the latency of the state transfer phase of migration. We suggest using a dedicated network for state transfer, with a nominal bandwidth of at least 1 Gbps.

2.7 AIX and active migration

An AIX partition continues running during an active migration. Most AIX features work seamlessly before, during, and after the migration. These include, but are not limited to, the following:

► System and advanced accounting

► Workload manager

► System trace

► Resource sets

Including exclusive use processor resource sets.

► Pinned memory

► Memory affinity

See 5.7, "Migration awareness" on page 151 for details on memory affinity considerations.

► Kernel and kernel extensions

See 5.9, "Making kernel extension migration aware" on page 160 for details on how to make a kernel extension migration aware.

► Large memory pages

Huge memory pages cannot be used

- ► Processor binding

  Processes remain bound to the same logical processor throughout the migration. See 5.7, "Migration awareness" on page 151 for details on processor binding considerations.

Performance monitoring tools (such as `topas`, `tprof`, `filemon`, and so on) can run on a mobile partition during an active migration. However, the data that these tools report during the migration process may not be significant, because of underlying hardware changes, performance monitor counters that may be reset, and so on.

Although AIX is migration safe, verify that any applications you are running are migration safe or aware. See 5.7, "Migration awareness" on page 151 for more information.

Details on Linux were not available at the time of publication.

3

Requirements and preparation

In this chapter, we discuss the different preparatory steps required for a logical partition to be migrated from one system to another system. There are a number of requirements to fulfill, and these depend on whether you want to perform an inactive or an active partition migration.

As already described in the first chapter, inactive partition migration lets you move a powered off logical partition, including its operating system and applications, from one system to another.

Active partition migration is the ability to move a running logical partition, including its operating system and applications, from one system to another without disrupting the operation of that logical partition.

We categorize and discuss these requirements in the following sections:

► HMC
► Systems
► Virtual I/O Servers
► Supported operating systems
► Storage
► Network

41

When you have ensured that all these requirements are satisfied and all preparation tasks are completed, the HMC verifies and validates the Live Partition Mobility environment. If this validation turns out to be successful, then you can initiate the partition migration by using the wizard on the HMC GUI or through the HMC CLI.

3.1 Requirements for Live Partition Mobility

The following are the major requirements for active Live Partition Mobility:

► Hardware Management Console requirements

The minimum requirements for the HMC version are release V7R320 or later for both active and inactive partition migration. If you are not at this level, you need to get the HMC machine code fixes and upgrades. The console requirements are at least at the 7310-CR2 or the 7310-C03 HMC.

► Source and destination system requirements

– The source and destination system must be an IBM System p POWER6-based model (like the IBM System p 570, also known as the 9117-MMA, which was used to produce this publication).

A system is capable of being either the source or destination of a migration if it contains the necessary processor hardware to support it. We will call this additional hardware capability *migration support*.

– Both the source and destination systems must be at firmware level eFW3.2 or later.

The current firmware level can be checked through the HMC after selecting the system, then **Update,** and finally **View system information**. If the version is not at the required level for Live Partition Mobility, you have to perform an update through the HMC by selecting **Upgrade Licensed Internal Code to a new release**.

Although there is a minimum required firmware level, each system may have a different level of firmware. The level of source system firmware must be compatible with the destination firmware.

► Source and destination Virtual I/O Server requirements

At least one Virtual I/O Server at release level 1.5 or higher has to be installed both on the source and destination systems.

A new partition attribute called the mover service partition has been defined that enables you to indicate whether a mover-capable Virtual I/O Server partition should be considered during the selection process of the MSP for a migration. By default, all Virtual I/O Server partitions have this new partition attribute set to FALSE.

In addition to having the mover partition attribute set to TRUE, a Virtual I/O Server must have a VASI device defined and configured in order to be considered a valid MSP candidate. A migration wizard on the HMC guides you through the configuration steps to help ensure that the system and the partitions are properly configured to support partition migration, which

includes designating one or more Virtual I/O Server partitions with the mover attribute and defining a VASI device in that partition.

The `ioslevel` command can be executed on a Virtual I/O Server in order to determine the current release of the Virtual I/O Server and to see if an upgrade is necessary.

► Operating system requirements

The operating system running in the mobile partition has to be AIX or Linux. A Virtual I/O Server logical partition or an i5/OS® logical partition cannot be migrated. The operating system must be at one of the following levels:

– AIX 5L™ Version 5.3 Technology Level 7 or later, AIX Version 6 or later.

– Red Hat Enterprise Linux Version 5 (RHEL5) Update 1 or later.

– SUSE Linux Enterprise Services 10 (SLES 10) Service Pack 1 or later.

Previous versions of AIX and Linux can participate in inactive partition migration if the operating systems support virtual devices and IBM System p POWER6-based systems.

► Storage requirements

For a list of supported disks and optical devices, see the data sheet available on the Virtual I/O Server support site at:

http://techsupport.services.ibm.com/server/vios/documentation/datash eet.html

► Network requirements

The migrating partition uses the virtual LAN for network access. The VLAN must be bridged (if there is more than one, then it also has to be bridged) to a physical network using a shared Ethernet adapter in the Virtual I/O Server partition. Your LAN must be configured such that migrating partitions can continue to communicate with other necessary clients and servers after a migration is completed.

3.2 Live Partition Mobility preparation checks

Table 3-1 gives you an overview of all the preparatory tasks required for a successful migration.

Table 3-1   Preparing the environment for Live Partition Mobility

Task	Details see page	Remarks for inactive migration
Preparing servers	46	
Preparing HMC	52	
Preparing VIOS	54	see Note 1
Preparing mobile partition	59	see Note 2
Preparing the storage	73	
Network considerations	80	

Table Note 1: For inactive migration, you have to perform fewer preparatory tasks on the Virtual I/O Server:

► You do not have to enable the mover service partition on either the source or destination Virtual I/O Server.

► You do not have to configure VASI on either the source or destination Virtual I/O Server.

► You do not have to synchronize the time-of-day clocks.

Table Note 2: For inactive migration, you have to perform fewer preparatory tasks on the mobile partition:

► RMC connections are not required.

► The mobile partition can have dedicated I/O. These dedicated I/O devices will be removed automatically from the partition before the migration occurs.

► Barrier Synchronization Registers can be used in the mobile partition.

► The mobile partition can use huge pages.

► The applications do not have to be migration-aware or migration-safe.

Some of the new settings can be changed dynamically (partition workload groups, mover service partitions, time-reference, and VASI), while others have to be changed statically (barrier synchronization registers and redundant error path reporting).

3.3 Preparing the systems for Live Partition Mobility

Careful planning in order to determine your environment is required before Live Partition Mobility can be successfully implemented. After the validation of all the required versions and levels in the previous section, this section describes the planning tasks that should be considered and completed on the source and destination systems before migrating a logical partition. These are required for inactive partition migration as well as active partition migration.

3.3.1 HMC

Ensure that the source and destination systems are managed by the same HMC (or a redundant HMC pair).

3.3.2 LMB size

Ensure that the logical memory block size is the same on the source and destination systems. The default LMB size depends on the amount of memory installed in the CEC. It varies between 16 MB and 256 MB. Figure 3-1 shows how the size of the logical memory block can be modified by using the Advanced System Management Interface (ASMI).
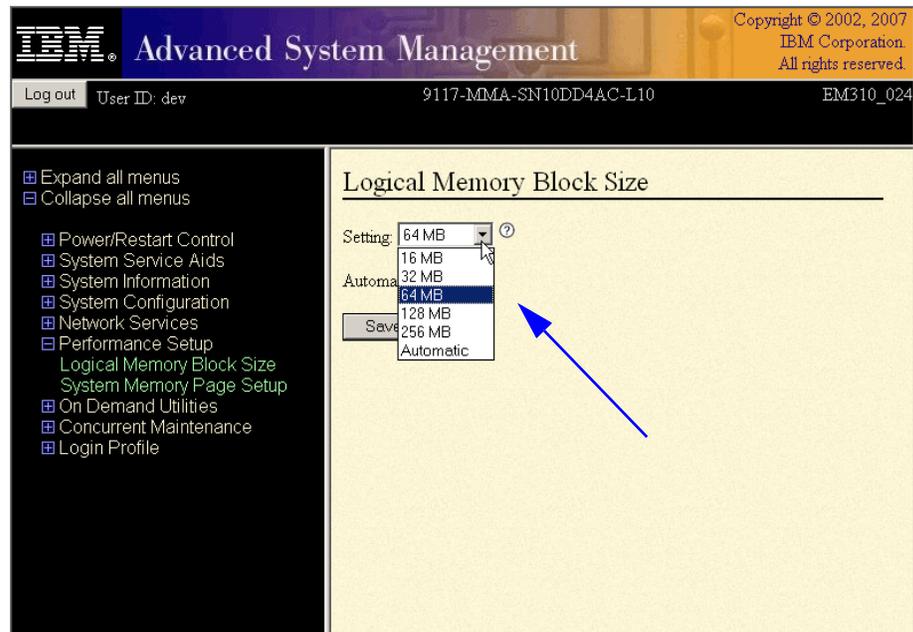
Figure 3-1   Checking and changing LMB size with ASMI

3.3.3 Battery power

Ensure that the destination system is not running on battery power. If the destination system is running on battery power, then you need to return the system to its regular power source before moving a logical partition to it. However, the source system can be running on battery power.

3.3.4 Available memory

Ensure that the destination system has enough available memory to support the mobile partition. To determine the available memory on the destination system and allocate more memory if necessary, you must have super administrator authority (a user with the HMC hscsuperadmin role, such as hscroot). The following steps have to be completed on the HMC:

1. You first have to determine the amount of memory of the mobile partition on the source system:

   a. In the navigation area, open **Systems Management**.

   b. Select the source system in the navigation area.

   c. In the contents area, select the mobile partition and select **Properties** in the task list.

   d. Select the **Hardware** tab and the **Memory** tab.

   e. View the **Memory** section and record the assigned memory settings.

   f. Click **OK**.

Figure 3-2 shows the result of the different actions.
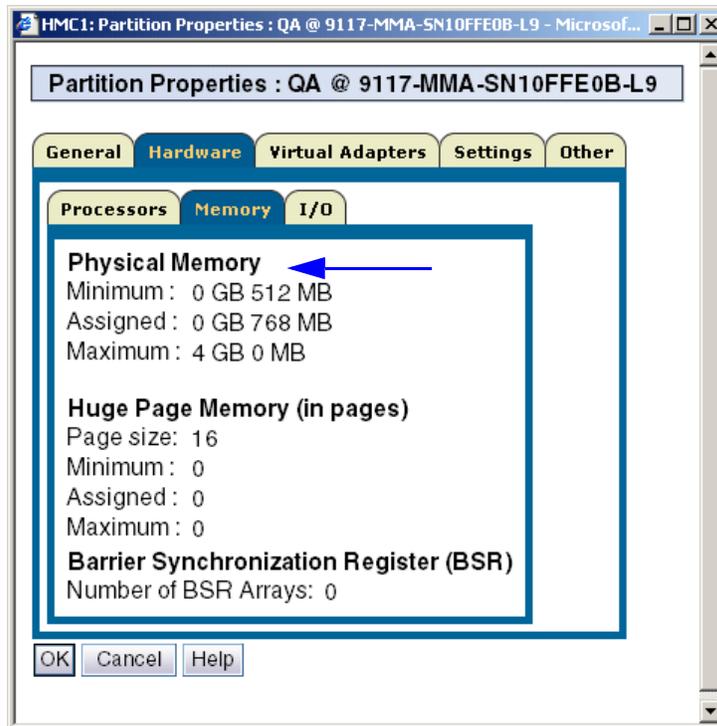
Figure 3-2   Checking the amount of memory of the mobile partition

2. Determine the memory available on the destination system:

    a. In the contents area, select the destination system and select **Properties** in the task list.

    b. Select the **Memory** tab.

    c. Record the Available memory and Current memory available for partition usage.

    d. Click **OK**.

3. Compare the values from the previous steps.

    – If the destination system has enough available memory to support the mobile partition, then skip the rest of this procedure and continue with other preparation tasks.

– If the destination system does not have enough available memory to support the mobile partition, you must dynamically free up some memory (or use the Capacity on Demand) feature, when available) on the destination system before the actual migration can take place.

3.3.5 Available processors to support Live Partition Mobility

Ensure that the destination system has enough available processors (or processing units) to support the mobile partition. The profile created on the destination server matches the source server's, therefore dedicated processors must be available on the target if that is what you are using, or enough processing units in a shared processor pool.

To determine the available processors on the destination system and allocate more processors if necessary, you must have super administrator authority (a user with the HMC hscsuperadmin role, such as hscroot). The following steps have to be completed on the HMC:

1. Determine how many processors the mobile partition requires:

   a. In the navigation area, open **Systems Management**.

   b. Select the source system in the navigation area.

   c. In the contents area, select the mobile partition and select **Properties** in the task list.

   d. Select the **Hardware** tab and the **Processors** tab.

   e. View the **Processor** section and record the processing units settings.

   f. Click **OK**.

Figure 3-3 shows the result of the different actions. Note that in more recent HMC levels, p6 may appear as POWER6.
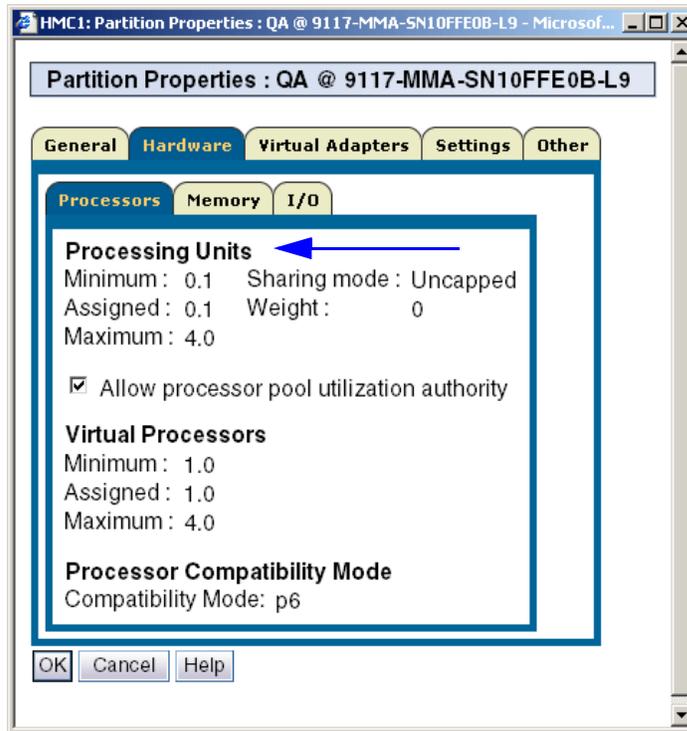
Figure 3-3   Checking the number of Processing Units of the mobile partition

2. Determine the processors available on the destination system:

   a. In the contents area, select the destination system and select **Properties** in the task list.

   b. Select the **Processors** tab.

   c. Record the Available processors available for partition usage.

   d. Click **OK**.

3. Compare the values from the previous steps.

   – If the destination system has enough available processors to support the mobile partition, then skip the rest of this procedure and continue with the remaining preparation tasks for Live Partition Mobility.

– If the destination system does not have enough available processors to support the mobile partition, you must dynamically free up some processors (or use the Capacity on Demand feature, when available) on the destination system before the actual migration can take place.

3.4 Preparing the HMC for Live Partition Mobility

The version and release of the HMC has to be at the correct level for Live Partition Mobility. See the requirements at the beginning of this chapter. Figure 3-4 shows how to check the current version and release of our HMC.
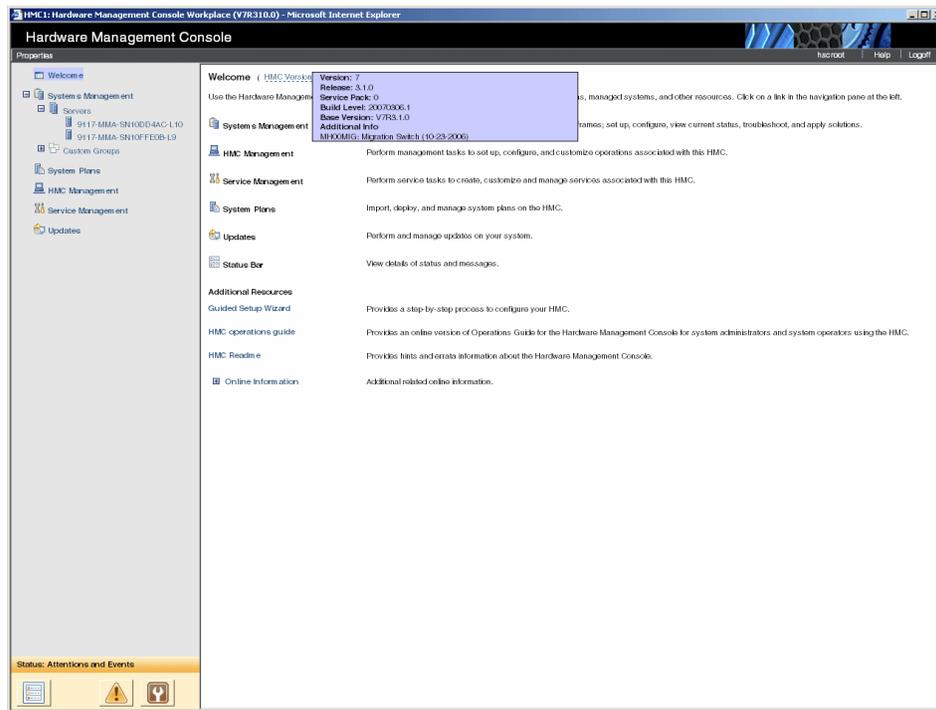
Figure 3-4   Checking the version and release of HMC

> **Note:** Live Partition Mobility requires HMC Version 7.3.2 or higher to be used. In this publication, we used an early development version of the HMC software, 7.3.1 (see Figure 3-4).

If the HMC is not at the correct version and release, an upgrade is required. See Figure 3-5 and Figure 3-6 on page 54. More information about upgrading the Hardware Management Console can be found at:

http://www14.software.ibm.com/webapp/set2/sas/f/hmc/home.html
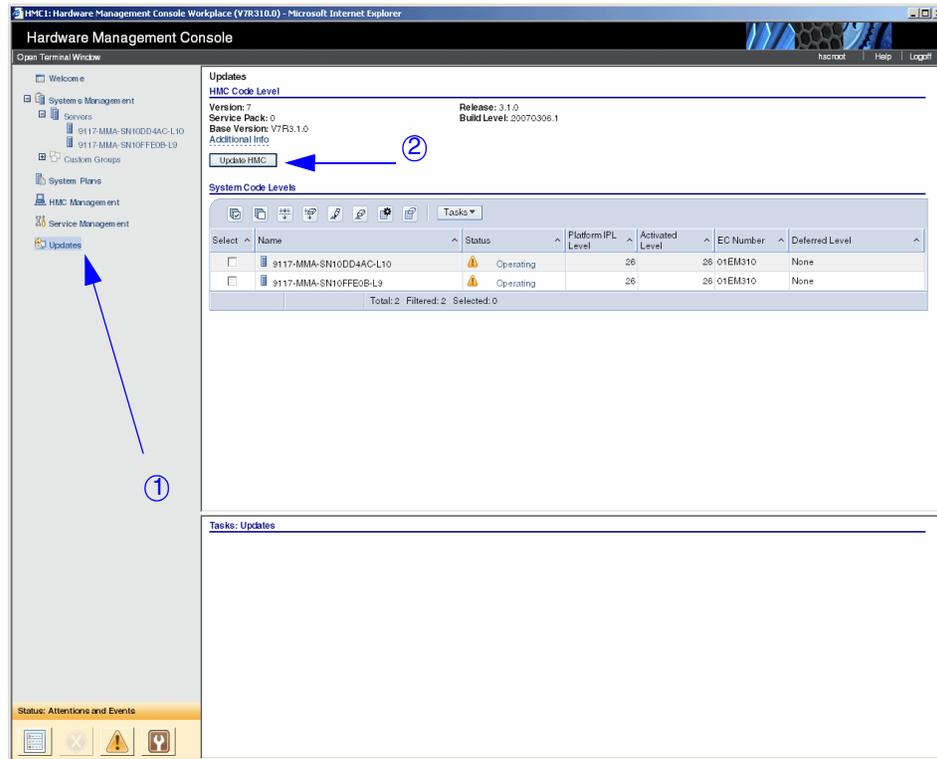
Figure 3-5   Upgrading the Hardware Management Console

Figure 3-6 appears when you click the **Update HMC** button (2) of Figure 3-5 on page 53.
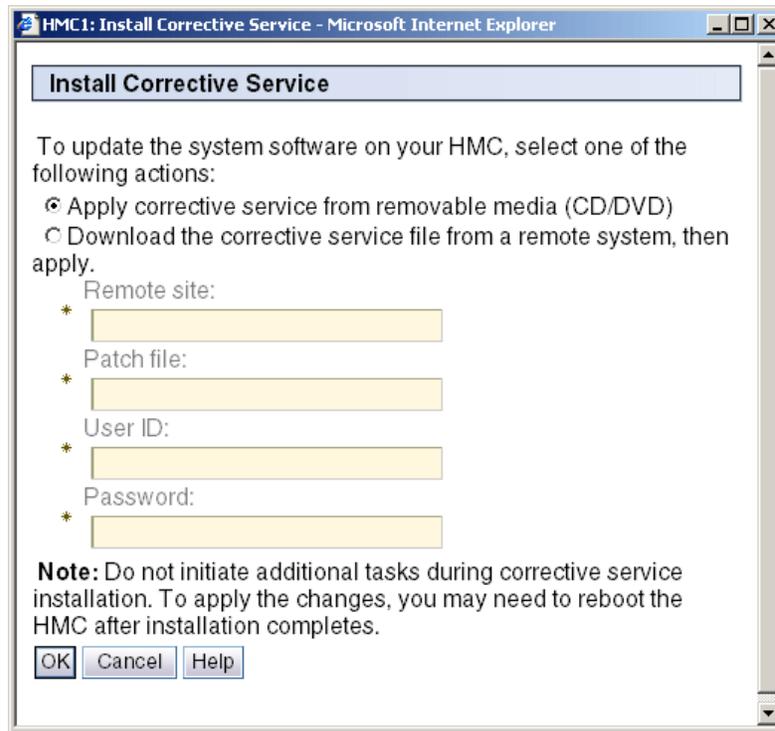
Figure 3-6   Install Corrective Service to upgrade the HMC

3.5 Preparing the Virtual I/O Servers

There are several tasks to complete in order to prepare the source and destination Virtual I/O Servers for Live Partition Mobility. There has to be at least one VIOS logical partition installed and activated on both the source and destination systems. For instructions on how to install the VIOS, see the following document on IBM InfoCenter:

http://publib.boulder.ibm.com/infocenter/eserver/v1r3s/topic/iphb1/iphb1.pdf

3.5.1 Virtual I/O Server version

Ensure that the source and destination Virtual I/O Servers are at Version 1.5 or higher. This can be checked on the Virtual I/O Server by running the following command:

```
ioslevel
```

If the source and destination Virtual I/O Servers do not meet the requirements, perform an upgrade.

3.5.2 Mover service partition

Ensure that at least one of the mover service partition (MSP) is enabled on a source and destination Virtual I/O Server partitions. The MSP is a Virtual I/O Server logical partition that has at least one VASI adapter configured to allow the MSP to communicate with the hypervisor.

There must be at least one MSP on both the source and destination Virtual I/O Servers for the mobile partition to participate in active partition migration. If the MSP is disabled on either the source or destination Virtual I/O Server, the mobile partition can be migrated inactively.

To enable the source and destination MSPs using the HMC, you must have super administrator (such as hscsuperadmin, as in the hscroot login) authority and complete the following steps:

1. In the navigation area, open **Systems Management** and select **Servers**.
2. In the contents area, open the source system.
3. Select the source Virtual I/O Server logical partition and select **Properties** on the task area.
4. On the **General** tab, select **Mover Service Partition**, and click **OK**.
5. Repeat these steps for the destination system.

Figure 3-7 shows the result of these actions.
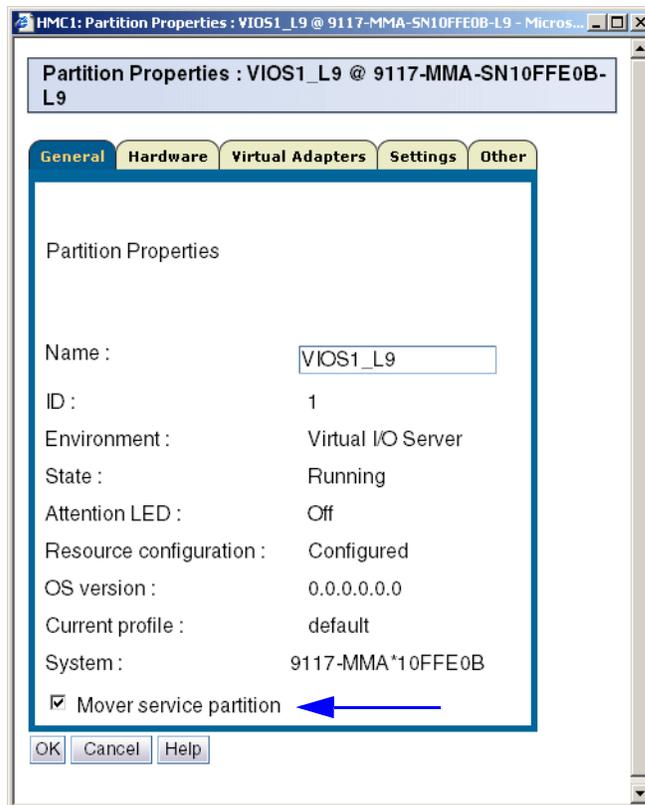
Figure 3-7   Enabling MSP

3.5.3 Configure VASI

Ensure that there is a VASI device configured on both the source and destination Virtual I/O Server partitions designated as mover service partitions. If the VASI device is not configured on either the source or destination Virtual I/O Server, the mobile partition can be migrated inactively.

The VASI device allows the mover service partition to communicate with the hypervisor and must be configured on both the source and destination Virtual I/O Servers designated as the mover service partitions for the mobile partition to participate in active partition migration.

> **Important:** As of the November 2007 release, configuring the VASI device is no longer required. The VASI devices is automatically configured by the HMC.

To configure a VASI device on the mover service partitions using the HMC, you must be a super administrator (hscroot) to complete the following steps:

1. In the navigation area, open **Systems Management** and select **Servers**.

2. In the contents area, open the source system.

3. Select the source Virtual I/O Server logical partition, select the active partition profile, and click **Properties**.

4. Select the **Virtual Adapters** tab and click **Actions** tab.

5. Click **Create** and select **VASI Adapter**.

6. The VASI device is placed in slot 2. Click **OK** in the pop-up window.

7. Click **OK** to exit the logical partition profile properties.

8. Repeat the previous steps for the destination system and the destination Virtual I/O Server.

Figure 3-8 shows the creation of a Virtual Asynchronous Service Interface (VASI).
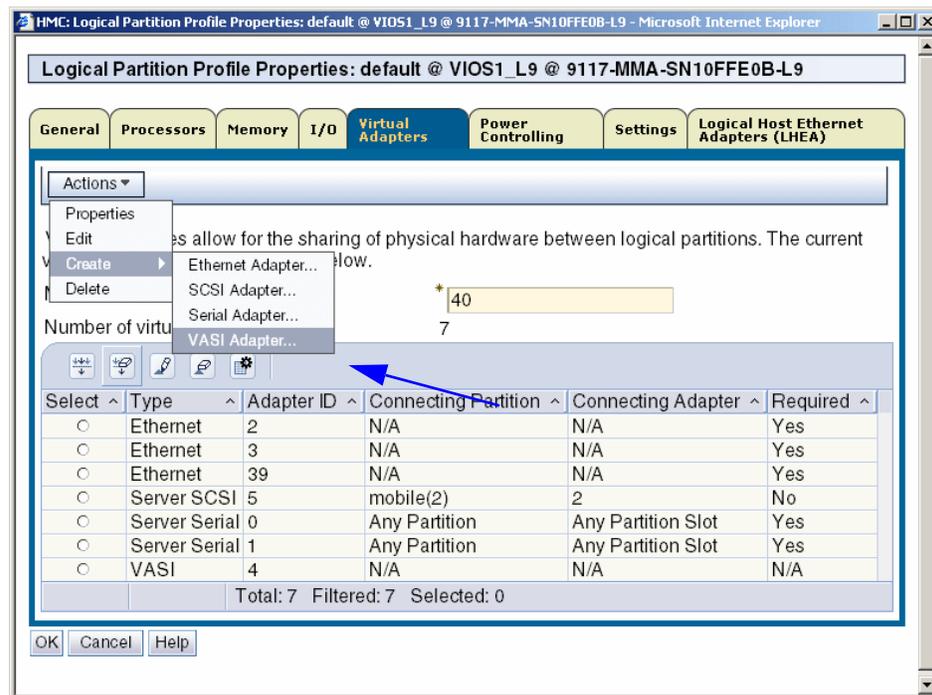
Figure 3-8   Configuring VASI

3.5.4 Synchronize time-of-day clocks

Another recommended, although optional, task for active partition migration is the synchronization of the time-of-day clocks for the source and destination Virtual I/O Server partitions.

If you choose not to complete this step, the source and destination Virtual I/O Servers will synchronize the clocks while the mobile partition is moving from the source system to the destination system. Completing this step before the mobile partition is moved can prevent possible errors.

To synchronize the time-of-day clocks on the source and destination Virtual I/O Servers using the HMC, you must be a super administrator (such as hscroot) to complete the following steps:

1. In the navigation area, open **Systems Management**.
2. Select **Servers** and select the source system.
3. In the contents area, select the source Virtual I/O Server logical partition.
4. Click on **Properties**.
5. Click the **Settings** tab.
6. For Time reference, select **Enabled** and click **OK**.
7. Repeat the previous steps for the destination system and the destination Virtual I/O Server.

Figure 3-9 shows the time-of-day synchronization.
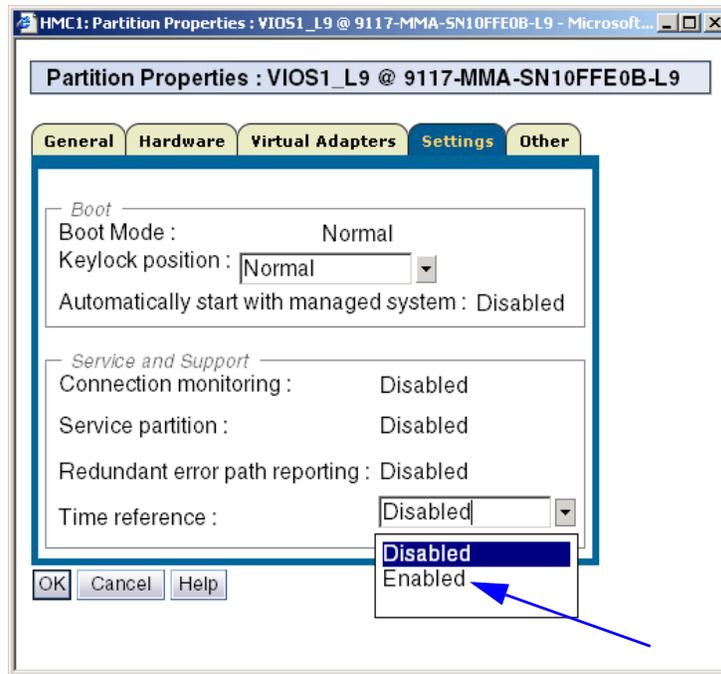
Figure 3-9   Synchronizing the time-of-day clocks

3.6 Preparing the mobile partition for mobility

This section describes the tasks that you must complete to prepare a mobile partition for Live Partition Mobility in order to have a successful migration.

3.6.1 Check operating system version

Ensure that the operating system meets the requirements for Live Partition Mobility. These requirements can be found in 3.1, "Requirements for Live Partition Mobility" on page 43.

3.6.2 RMC

Ensure that for active partition migration that Resource Monitoring and Control (RMC) connections are established.

RMC can be configured to monitor resources and perform an action in response to a defined condition. The flexibility of RMC enables you to configure response actions or scripts that manage general system conditions with little or no involvement from the system administrator.

To establish a RMC connection for the mobile partition, you must be a super administrator (a user with the HMC hscsuperadmin role, such as hscroot) on the HMC and complete the following steps:

1. Sign on to the operating system with root authority.

2. From the command line, enter the following command:

   `lsrsrc IBM.ManagementServer`

   – If the command output includes `ManageType = "HMC"`, then the RMC connection is established and you can continue with the additional preparation tasks

   – If you received a message indicating that there is no IBM.ManagementServer resource or ManagerType does not equal HMC, then continue to the next step.

3. Establish the RMC connection specifically for your operating system. For Linux, this means installing RSCT utilities.

   – AIX: See *Configuring Resource Monitoring and Control (RMC) for the Partition Load Manager*, found at:

   http://publib.boulder.ibm.com/infocenter/eserver/v1r3s/topic/iphb k/iphbkrmc_configuration.htm

   – Red Hat Linux: Additional software (RSCT Utilities) for Red Hat on HMC managed servers.

   – SUSE Linux: Installing additional software (RSCT Utilities) for SUSE Linux on HMC managed servers.

   – For Linux, these tools can be downloaded from (take the HMC- or IVM-managed server link):

   http://www14.software.ibm.com/webapp/set2/sas/f/lopdiags/home.htm l

3.6.3 Disable redundant error path reporting

Ensure that the mobile partition is not enabled for redundant error path reporting.

Redundant error path reporting allows a logical partition to report server common hardware errors and partition hardware errors to the HMC. Redundant error path reporting should be disabled if you want to migrate a logical partition.

To disable redundant error path reporting for the mobile partition, you must be a super administrator and complete the following steps:

1. In the navigation area, open **Systems Management**.

2. Select **Servers** and select the source system.

3. In the contents area, select the logical partition you wish to migrate and select **Configuration** → **Manage Profiles**.

4. Select the partition profile and click **Properties** (Later HMC levels have this function in the Action pull-down menu and the Edit action item).

5. Click the **Settings** tab.

6. Deselect **Enable redundant error path reporting**, and click **OK**.

7. Because disabling redundant error path reporting cannot be done dynamically, you have to shut down the mobile partition, then power it on using the profile with the modifications.

Figure 3-10 shows the disabled redundant error path handling:
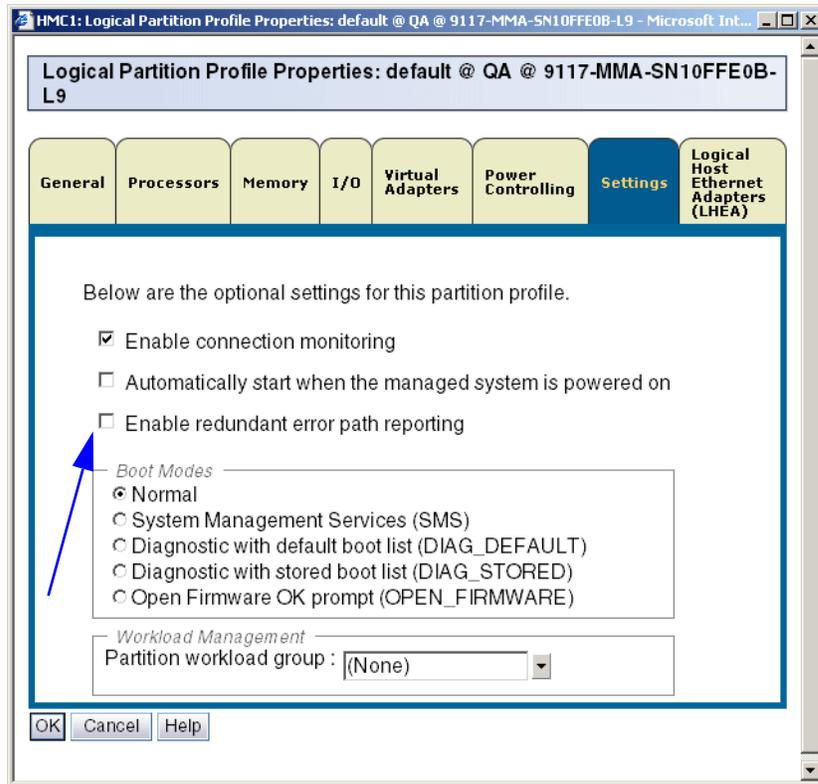
Figure 3-10   Disable redundant error path handling

3.6.4 Virtual serial adapters

Ensure that the mobile partition is not using a virtual serial adapter.

Virtual serial adapters are often used for virtual terminal connections to the operating system. The first two virtual serial adapters (slots 0 and 1) are reserved for the Hardware Management Console (HMC). For a logical partition to participate in a partition migration, it cannot have any required virtual serial adapters, except for the two reserved for the HMC.

To dynamically disable unreserved virtual serial adapters using the HMC, you must be a super administrator and complete the following steps:

1. In the navigation area, open **Systems Management**.

2. Select **Servers** and select the source system.

3. In the contents area, select the logical partition you wish to migrate and select **Configuration** → **Manage Profiles**.

4. Select the partition profile and click **Properties**.

5. Select the **Virtual Adapter** tab.

6. Expand **Virtual Serial**.

7. If there are more than two virtual serial adapters listed, then ensure that the adapters in slots 2 and higher are not selected as Required.

8. Click **OK**.

3.6.5 Partition workload groups

Ensure that the mobile partition is not part of a logical partition group.

A partition workload group identifies a set of partitions that reside on the same system. The partition profile specifies the name of the partition workload group to which it belongs, if applicable. For a logical partition to participate in a partition migration, it cannot be assigned to a partition workload group.

To dynamically remove the mobile partition from a partition workload group, you must be a super administrator on the HMC and complete the following steps:

1. In the navigation area, open **Systems Management**.

2. Select **Servers**.

3. In the contents area, open the source system.

4. Select the mobile partition and select **Properties**.

5. Click the **Other** tab.

6. In the **New partition workload group field**, select **(None)**.

7. In the contents area, open the mobile partition and select **Configuration** → **Manage Profiles**.

8. Select the active logical partition profile and click **Properties**.

9. Click the **Settings** tab.

10. In the **Workload Management** area, select **(None)** and click **OK**.

11. Repeat the last three steps for all partition profiles associated with the mobile partition.

Figure 3-11 and Figure 3-12 on page 65 show the different tabs for the disablement of the partition workload group (both in the partition as well as in the partition profiles).
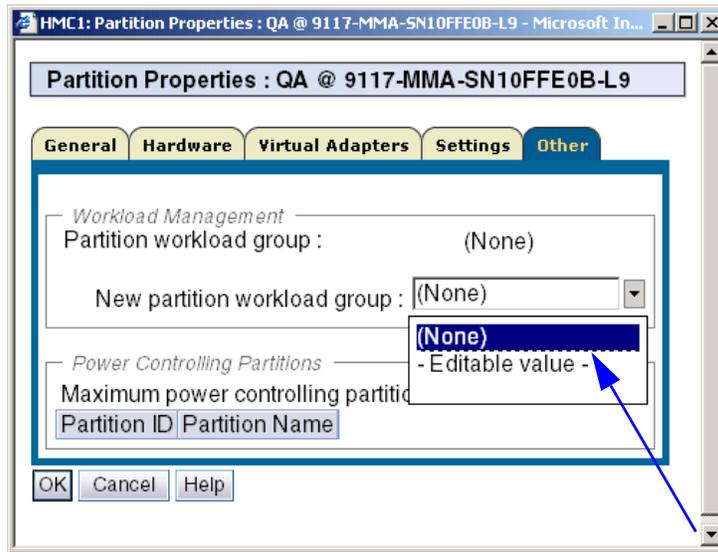
Figure 3-11   Disabling partition workload group - Other tab
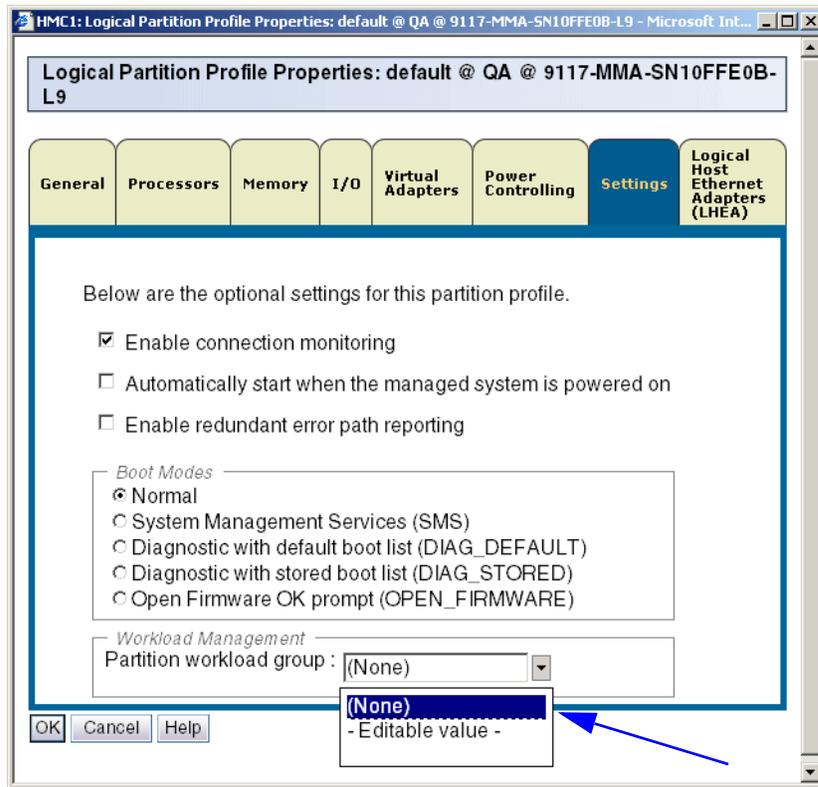
Figure 3-12   Disabling partition workload group - Settings tab

3.6.6 Barrier Synchronization Register

Ensure that the mobile partition is not using barrier synchronization register (BSR) arrays.

BSR is a memory register that is located on certain POWER- based processors. A parallel-processing application running on AIX can use a BSR to perform barrier synchronization, which is a method for synchronizing the threads in the parallel-processing application. For a logical partition to participate in active partition migration, it cannot use BSR arrays. However, it can still participate in inactive partition migration if it uses BSR.

To disable BSR for the mobile partition using the HMC, you must be a super administrator and complete the following steps:

1. Open the source system.

2. Open **Partitions**.

3. Select the mobile partition and select **Properties**.

4. Click the **Hardware** tab.

5. Click the **Memory** tab.

6. If the number of BSR arrays equals zero, the mobile partition can participate in inactive or active migration. This is shown on Figure 3-13. We can now continue with additional preparatory tasks for the mobile partition.
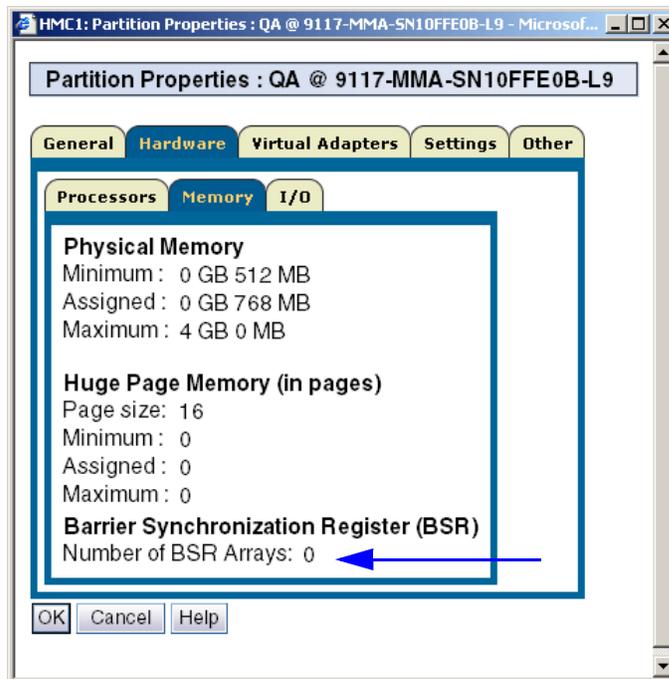
Figure 3-13   Checking the number of BSR arrays on the mobile partition

7. If the number of BSR arrays is not equal to zero, then take one of the following actions:

   – Perform an inactive migration instead of an active migration.

   – Click **OK** and continue to the next step to prepare the mobile partition for an active migration.

8. In the contents area, open the mobile partition and select **Configuration** → **Manage Profiles**.

9. Select the active logical partition profile and click **Properties**.

10. Click the **Memory** tab.

11. Enter 0 in the BSR arrays for this profile field and click **OK**. This is shown in Figure 3-14.

12. Because modifying BSR cannot be done dynamically, you have to shut down the mobile partition, then power it on using the profile with the BSR modifications.
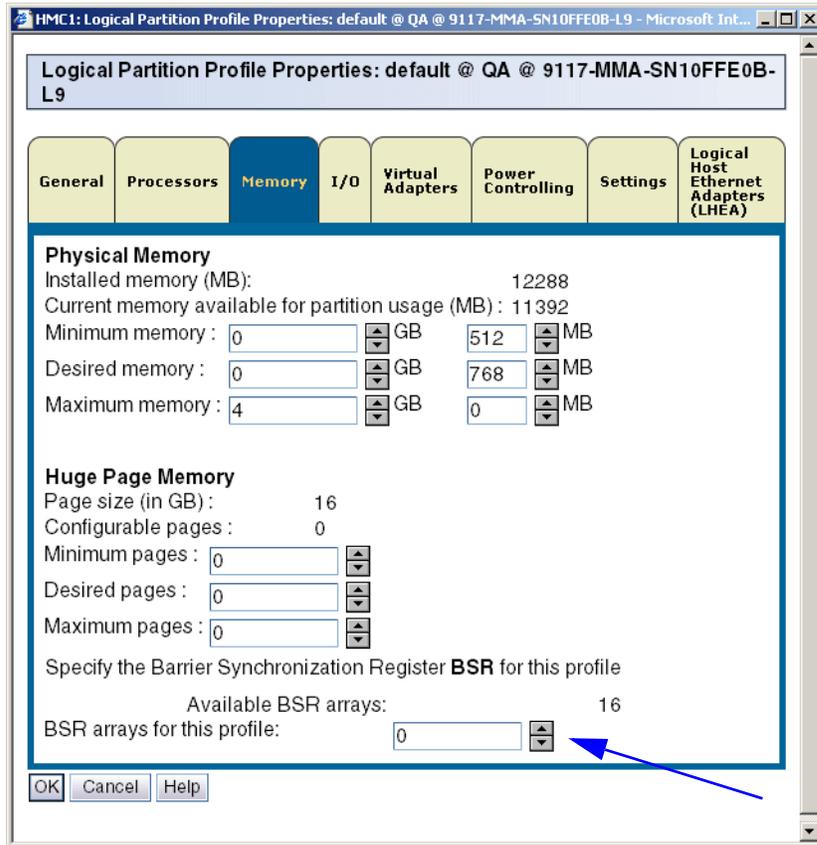
Figure 3-14   Setting number of BSR arrays to zero

3.6.7 Huge pages

Ensure that the mobile partition is not using huge pages.

Huge pages can improve performance in specific environments that require a high degree of parallelism, such as in DB2® partitioned database environments. You can specify the minimum, desired, and maximum number of huge pages to assign to a partition when you create a partition profile. For a logical partition to participate in active partition migration, it cannot use huge pages. However, if the mobile partition does use huge pages, it can still participate in inactive partition migration.

To configure huge pages for the mobile partition using the HMC, you must be a super administrator and complete the following steps:

1. Open the source system and select **Properties**.

2. Click the **Advanced** tab.

- If the current huge page memory equals 0, as you can see in Figure 3-15, then skip the rest of this procedure and continue with additional preparatory tasks for the mobile partition.
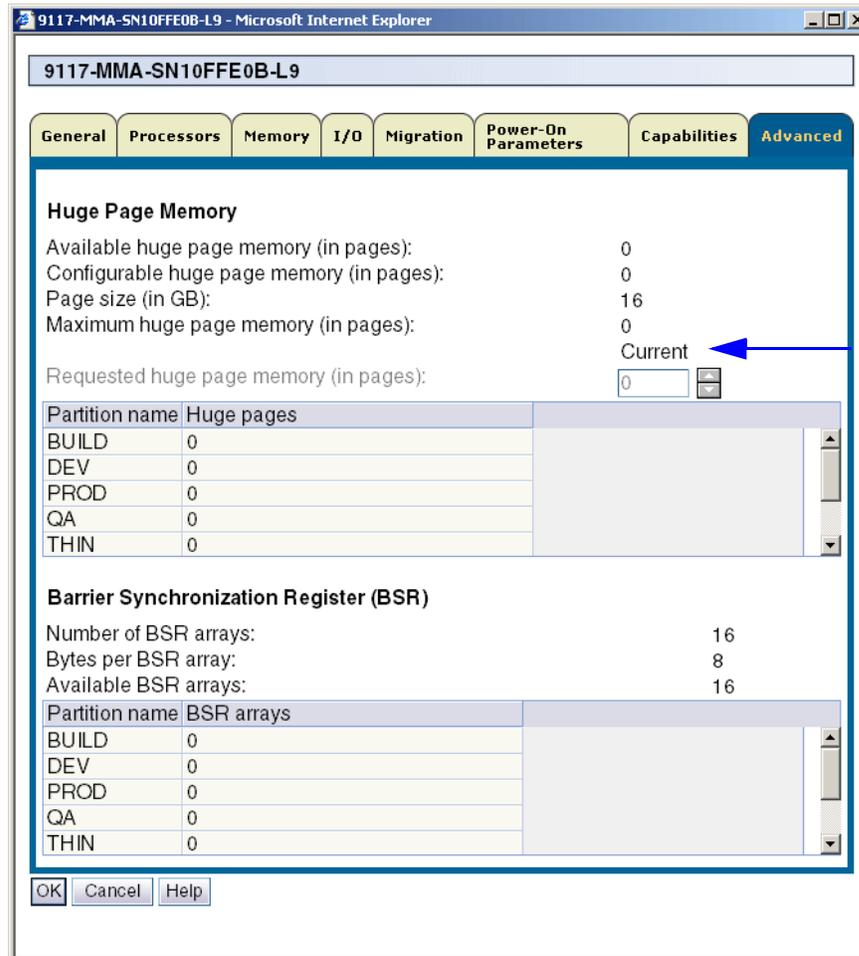
Figure 3-15   Checking if Huge Page memory equals zero

- If the current huge page memory is not equal to 0, then take one of the following actions:

  - Perform an inactive migration instead of an active migration.

  - Click **OK** and continue with the next step to prepare the mobile partition for an active migration.

3. In the contents area, open the mobile partition and select **Configuration** → **Manage Profiles**.

4. Select the active logical partition profile and click **Properties**.

5. Click the **Memory** tab.

6. Enter 0 in the field for desired huge page memory, and click **OK**. This is shown in Figure 3-16.

7. Shut down the mobile partition, then power it on using the profile with the huge page modifications.
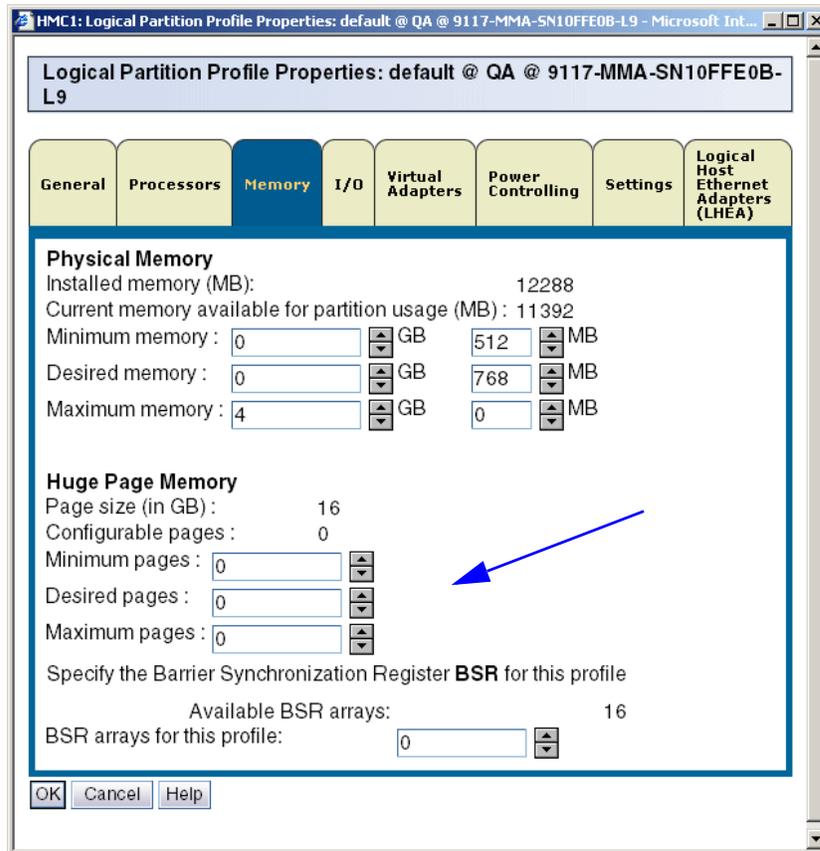
Figure 3-16   Setting Huge Page Memory to zero

3.6.8 Physical - dedicated I/O

Ensure that the mobile partition does not have physical or dedicated (required) I/O adapters and devices.

For a logical partition to participate in active partition migration, it cannot have any required or physical I/O. All I/O must be virtual. If the mobile partition has required or physical I/O, it can participate in inactive partition migration. After migration, the required or physical I/O configuration must be verified. Physical desired I/O can be removed dynamically with a dynamic LPAR operation.

To remove required I/O from the mobile partition using the HMC, you must be a super administrator and complete the following steps:

1. In the navigation area, open **Systems Management**.

2. Select **Server** and select the source system.

3. In the contents area, open the mobile partition and select **Configuration** → **Manage Profiles**.

4. Select the active logical partition profile and click **Properties**.

5. Click the **I/O** tab.

    a. If Required is not selected for any resource, as shown in Figure 3-17, then skip the rest of this procedure and continue with additional preparatory tasks for the mobile partition.
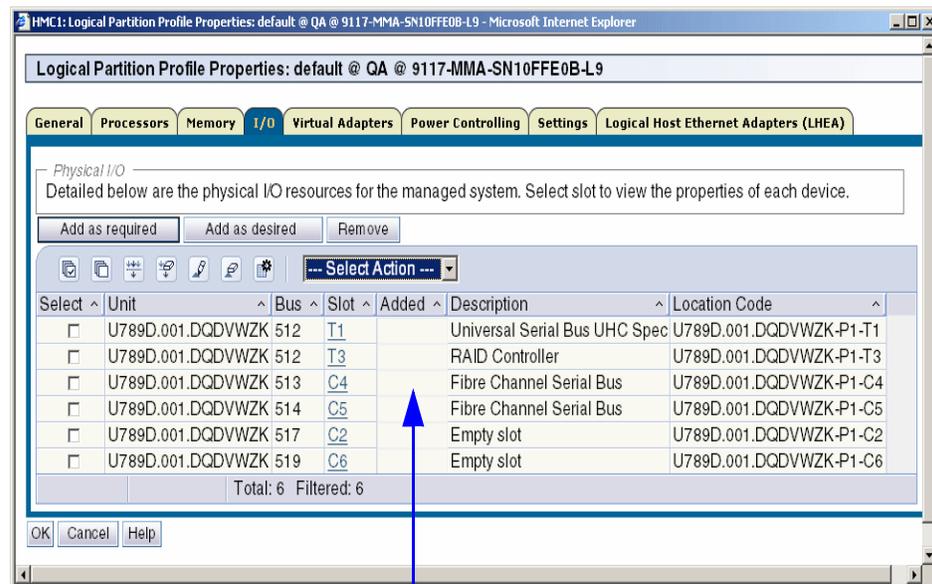

Figure 3-17   Checking if there are required resources in the mobile partition

b. If Required is selected for any resource, take one of the following actions:

- Perform an inactive migration instead of an active migration.
- Continue with the next step to prepare the mobile partition for an active migration.

6. For each resource that is selected as Required, deselect **Required** and click **OK**.

7. Shut down the mobile partition, then power it on using the profile with the required I/O resource modifications.

> **Note:** You must also verify that no HEA devices are configured, because these are also considered as physical I/O. Inactive migration is still possible if LHEA are configured.

Figure 3-18 shows you how to verify if an HEA is configured for the mobile partition. We first have to select an HEA in order to define an LHEA(1), and then verify if there are logical port IDs (2). If there is no logical port ID in this column, then there is no logical host Ethernet adapter configured for this partition. More information about HEA can be found on page 171.

Figure 3-18   Logical Host Ethernet Adapter

3.6.9 Name of logical partition profile

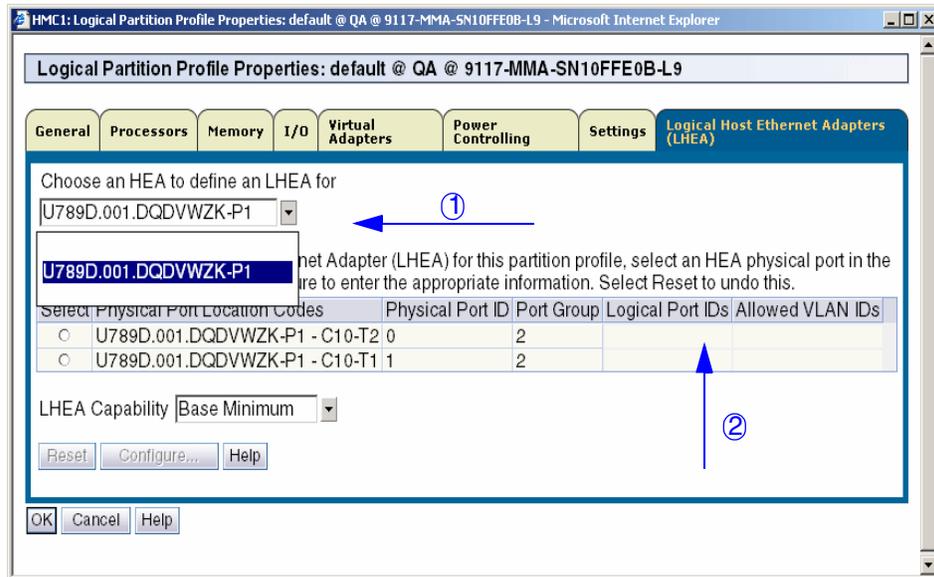Determine the name of the logical partition profile for the mobile partition on the destination system. This is an optional step. As part of the migration process, the HMC will create a new migration profile containing the partition's current state. Unless you specify a profile name when you start the migration, this profile will replace the existing profile that was last used to activate the partition. Also, if you specify an existing profile name, the HMC will replace that profile with the new migration profile. If you do not want the migration profile to replace any of the partition's existing profiles, you must specify a unique profile name. The new profile contains the partition's current configuration and any changes that are made during the migration.

3.6.10 Mobility-safe or mobility-aware

Ensure that the applications running in the mobile partition are mobility-safe or mobility-aware. See 5.7, "Migration awareness" on page 151 and 5.8, "Making applications migration-aware" on page 153for more information.

3.6.11 Changing partition profiles

If you changed any partition profile attributes, you need to shut down and activate the new profile in order for the new values to take effect. Use the following tasks to complete these steps.

1. In the contents area, select the mobile partition, click **Operations**, and click **Shut down**.

2. In the contents area, select the mobile partition, click **Operations**, click **Activate**, and select the logical partition profile.

3.7 Configuring the external storage

This section describes the tasks that you must complete to ensure your storage configuration meets the minimal configuration for Live Partition Mobility before you can actually migrate your logical partition.

1. Verify that the same SAN disks used as virtual disks by the mobile partition are assigned to the source and destination Virtual I/O Server logical partitions.

2. Verify, with the **lsdev** command, that the reserve_policy attributes on the shared physical volumes are set to no_reserve on the Virtual I/O Servers.

   To list all the disks, type the following command:

   ```
   lsdev -type disk
   ```

   To list the attributes of hdisk*X*, type the following command:

   ```
   lsdev -dev hdiskX -attr
   ```

   If reserve_policy is not set to no_reserve, use the following command:

   ```
   chdev -dev hdiskX -attr reserve_policy=no_reserve
   ```

3. Verify that the physical volume has the same unique identifier, physical identifier, or an IEEE volume attribute. These identifiers are required in order to export a physical volume as a virtual device.

   – To list disks with a unique identifier (UDID), you must complete the following steps:

     i. Type the **oem_setup_env** command on the VIOS CLI.

     ii. Type the **odmget -qattribute=unique_id CuAt** command in order to list the disks that have a UDID.

Example 3-1   Output of odmget command

```
CuAt:
    name = "hdisk1"
    attribute = "unique_id"
    value = "200B75AMYW1100107210790003IBMfcp"
    type = "R"
    generic = ""
    rep = "nl"
    nls_index = 79

CuAt:
    name = "hdisk2"
    attribute = "unique_id"
    value = "200B75AMYW1100207210790003IBMfcp"
    type = "R"
    generic = ""
    rep = "nl"
    nls_index = 79
```

     iii. Type **exit** to return to the VIOS prompt.

– To list disks with a physical identifier (PVID), you must complete the following steps:

    i. Type the **lspv** command to list the devices with a PVID. If the second column has a value of none, the physical volume does not have a PVID. You have to put a PVID on the physical volume before it can be exported as a virtual device.

Example 3-2   Output of lspv command

```
$ lspv
NAME            PVID                            VG            STATUS
hdisk0          00cdd4ac470c687e                rootvg        active
hdisk3          00cdd4ac4d617bcd                None
hdisk4          00cdd4ac5276f2fe                None
```

    ii. Type the **chdev** command to put a PVID on the physical volume:

```
chdev - dev physicalvolumename -attr pv=yes -perm
```

– To list disks with an IEEE volume attribute identifier, you must issue (in **oem_setup_env)**the following command:

```
lsattr -El hdiskX
```

4. Verify that the mobile partition has access to a source Virtual I/O Server virtual SCSI adapter. You need to verify the configuration of the virtual SCSI adapters on the mobile partition and the source Virtual I/O Server logical partition in order to ensure that the mobile partition has access to storage. To do so, you must be a super administrator (such as hscroot) to complete the following steps:

a. Verify the virtual SCSI adapter configuration of the mobile partition:

    i. In the navigation area, open **Systems Management**.

    ii. Click **Servers**.

    iii. In the contents area, open the source system.

    iv. Open **Partitions**.

    v. Select the mobile partition and click **Properties**.

    vi. Click the **Virtual Adapters** tab.

    vii. Select the **Virtual SCSI** tab.

    viii. Record the Slot ID and the Remote Slot ID for each virtual SCSI adapter.

    ix. Click **OK**.

b. Verify the virtual SCSI adapter configuration of the source Virtual I/O Server virtual SCSI adapter:

    i. In the navigation area, open **Systems Management**.

    ii. Click **Servers**.

    iii. In the contents area, open the source system.

    iv. Open **Partitions**.

    i. Select the Virtual I/O Server logical partition and click **Properties**.

    ii. Click the **Virtual Adapters** tab.

    iii. Select the **SCSI** tab.

    iv. Verify the following information:

       • The Slot ID corresponds to the Remote Slot ID that you recorded in "Record the Slot ID and the Remote Slot ID for each virtual SCSI adapter." on page 75 for the virtual SCSI adapter on the mobile partition.

       • The Remote Slot ID is either blank or corresponds to the Slot ID that you recorded in "Record the Slot ID and the Remote Slot ID for each virtual SCSI adapter." on page 75 for the virtual SCSI adapter on the mobile partition.

    v. Click **OK**.

c. If the values are incorrect, then complete the following steps:

    i. Plan the slot assignments and connection specifications for the virtual SCSI adapters by using a Virtual I/O adapter worksheet similar to Table 3-2.

Table 3-2   Virtual SCSI adapter worksheet

Virtual SCSI Adapter	Slot number	Connection specification
Source Virtual I/O Server virtual SCSI adapter		
Destination Virtual I/O Server virtual SCSI adapter		
Mobile partition Virtual SCSI adapter on source system		
Mobile partition Virtual SCSI adapter on destination system		

ii. When all virtual SCSI adapters on the source Virtual I/O Server logical partition allow access to virtual SCSI adapters of every logical partition (not only the mobile partition), you have two solutions:

- You can create a new virtual SCSI server adapter on the source Virtual I/O Server and allow only the virtual SCSI client adapter on the mobile partition to access it.

- You can change the connection specifications of a virtual SCSI server adapter on the source Virtual I/O Server so that it allows access to the virtual SCSI adapter on the mobile partition. This means that the virtual SCSI adapter of the client logical partition that currently has access to the virtual SCSI adapter on the source Virtual I/O Server will no longer have access to the adapter.

5. Verify that the destination Virtual I/O Server has sufficient free virtual slots to create the virtual SCSI adapters required to host the mobile partition in order to create a virtual SCSI adapter after it moves to the destination system. To verify the virtual SCSI configuration using the HMC, you must be a super administrator (such as hscroot) to complete the following steps:

a. Verify the virtual slot configuration of the destination Virtual I/O Server:

b. In the navigation area, open **Systems Management**.

c. Select **Servers**.

d. In the contents area, open the destination system.

e. Select the destination Virtual I/O Server logical partition and click **Properties**.

f. Select the **Virtual Adapters tab** and compare the number of virtual adapter to the maximum virtual adapters. This is shown in Figure 3-19,
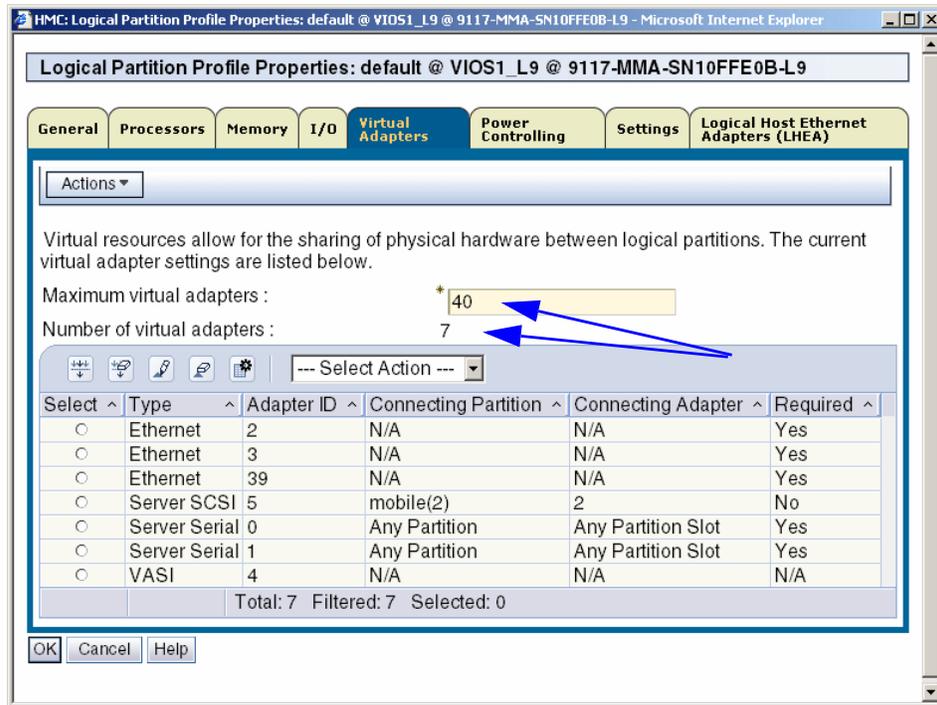

Figure 3-19   Checking free virtual slots.

– If, after verification, the number of maximum virtual adapters is higher or equal to the number of virtual adapters plus the number of virtual SCSI adapters required to host the migrating partition, you can continue with additional preparatory tasks.

– If the maximum virtual adapter value does not allow the addition of required virtual SCSI adapters for the mobile partition, then you have to modify its partition profile.

Click **System Management**, expand **Servers,** select the destination system, select the destination Virtual I/O Server, click in the task area on configuration, click **Manage profiles**, select the profile, and then click **Properties**.

Click the **Virtual Adapters** tab and modify (increase) the number of virtual adapters.

6. Verify that the mobile partition has access to the same physical storage on the storage area network from both the source and destination environments. This requirement has to be fulfilled for Live Partition Mobility to be successful.

   a. In the source environment, this means that the following connections must exist:

      i. A virtual SCSI client adapter on the mobile partition must have access to a virtual SCSI adapter on the source Virtual I/O Server logical partition.

      ii. That virtual SCSI server adapter on the source Virtual I/O Server logical partition must have access to a remote storage adapter on the source Virtual I/O Server logical partition.

      iii. That remote storage adapter on the source Virtual I/O Server logical partition must be connected to a storage area network and have access to some physical storage on the network.

   b. In the destination environment, this means that the following connection must exist:

      i. A remote storage adapter on the destination Virtual I/O Server logical partition has access to the same physical storage as the source Virtual I/O Server logical partition.

   To verify the virtual adapter connections using the HMC, you must be a super administrator and complete the following steps:

   a. Select **Systems Management.**

   b. Select **Servers**.

   c. In the contents area, open the source system and select a mobile partition.

   d. Select the mobile partition, select **Hardware (information)**, select **Virtual I/O adapter**, and select **SCSI**.

   e. Verify all the information and click **OK**.

      • If the information is correct, go to the next step.

      • If the information is incorrect, you have to return to the beginning of this section and complete the task associated with the incorrect information.

   f. In the contents area, open the destination system.

   g. Open **Partitions**.

   h. Select the destination Virtual I/O Server logical partition, select **Hardware (information)**, select **Virtual I/O adapter**, and select **SCSI**.

   i. Verify the information and click **OK**.

7. Verify that the mobile partition does not have physical or required I/O adapters and devices. This is only an issue for active partition migration. If you want to perform an active migration, you must move the physical or required I/O from the mobile partition, as explained in 3.6.8, "Physical - dedicated I/O" on page 71.

8. All profile changes on the mobile partition's profile must be activated before starting the migration in order for the new values to take effect.

    a. If the partition is not activated, it must be powered on. It is sufficient to activate the partition to the SMS menu.

    b. If the partition is active, you can shut it down and power on the partition again using the changed logical partition profile.

3.8 Network considerations

You must prepare and configure the network for partition migration. There are several tasks that you must complete to ensure your network configuration meets the minimal configuration for Live Partition Mobility.

► You first have to create a shared Ethernet adapter on the Virtual I/O Server using the HMC so that the client logical partitions can access the external network without needing a physical Ethernet adapter. Shared Ethernet adapters are required on both source and destination Virtual I/O Servers for all the external networks used by mobile partitions. If you plan to use a shared Ethernet adapter (SEA) with a host Ethernet adapter, ensure that the host Ethernet adapter (HEA) is set to promiscuous mode for the Virtual I/O Server. More information about HEA can be found in "Integrated Virtual Ethernet adapter overview" on page 172. If in promiscuous mode, it can only be used by a single LPAR.

The following steps have to be performed on the source and destination Virtual I/O Servers:

> **Note:** Link Aggregation or EtherChannel can also be used as the shared Ethernet adapter.

> **Note:** If you plan to use the host Ethernet adapter with the shared Ethernet adapter, ensure that you use the logical host Ethernet adapter to create the shared Ethernet adapter.

1. Ensure that you connect the source and destination Virtual I/O Servers and the shared Ethernet adapter to the network.

2. Configure virtual Ethernet adapters for the source and destination Virtual I/O Server partitions.

3. Ensure that the mobile partition has a virtual Ethernet adapter created by using the HMC GUI.

4. Activate the mobile partition to establish communication between its virtual Ethernet and the Virtual I/O Servers virtual Ethernet.

5. Verify that the operating system on the mobile partition sees the new Ethernet adapter by using the following command:

   ```
   lsdev -Cc adapter
   ```

6. The final step is to check that the client partition can access the external network. In order to do so, you must configure the TCP/IP connections for the virtual adapters on the client logical partitions by using the client partitions' operating systems (AIX or Linux).

   ```
   mktcpip -h hostname -a IPaddress -i interface -g gateway
   ```

3.9 Distance considerations

There are no architected maximum distances between systems for Live Partition Mobility. The maximum distance is dictated by the network and storage configuration used by the systems, and by the fact that both systems must be managed by the same HMC. Provided both systems are on the same network, are connected to the same shared storage, and are managed by the same HMC, then Live Partition Mobility will work. Standard long-range network and storage performance considerations apply.

3.10 Skill considerations

Live Partition Mobility builds on top of several existing technologies. Familiarity with them is helpful when working with Live Partition Mobility. This publication assumes a working knowledge of the following topics:

► Advanced POWER Virtualization

  – Virtual I/O Server.

  – Virtual SCSI.

  – Virtual and shared Ethernet.

  – See *Advanced POWER Virtualization on IBM System p5: Introduction and Configuration*, SG24-7940.

- ► Hardware Management Console
- ► Storage area networks
  - – Configuring shared storage is required for Live Partition Mobility.
- ► Dynamic logical partitioning
- ► AIX or Linux

4

Basic partition migration scenario

This chapter introduces the basics of configuring a Live Partition Mobility environment on IBM System p servers, and shows detailed steps of how to migrate a logical partition from a source system to a destination system in a single flow.

The chapter contains the following:

4.1 Basic Live Partition Mobility environment

This section shows you a simple configuration for Live Partition Mobility. It is composed of two systems. A single Virtual I/O Server partition is configured on each system. Using this configuration (Figure 4-1), we will explain the basic components involved in the Live Partition Mobility and will guide you through the configuration and execution tasks.
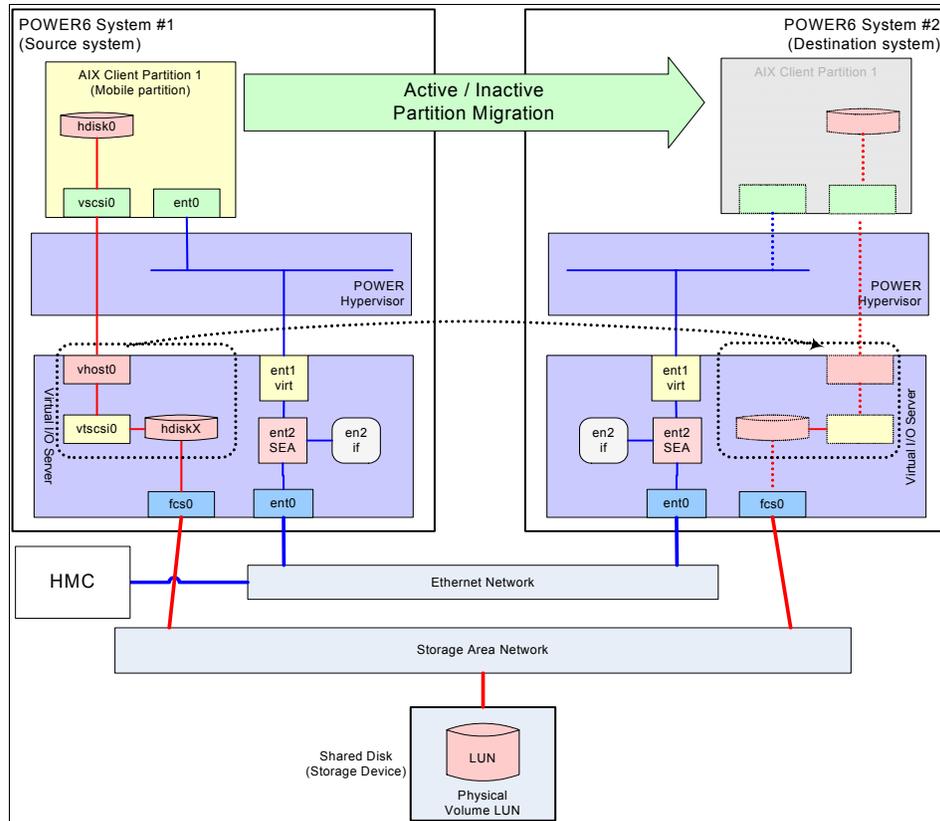

Figure 4-1   Basic Live Partition Mobility configuration

The following are the minimal requirements to configure the basic Live Partition Mobility environment.

► Hardware Management Console

   Both the source and the destination systems must be managed by the same Hardware Management Console (HMC).

► Same logical memory block size

The logical memory block size must be the same on the source and the destination system. You can check and update the logical memory block (LMB) size by using the Advanced System Management Interface. Refer to 3.3.2, "LMB size" on page 47 on how to change the LMB size.

► Mobile partition

A logical partition that is migrated from the source to the destination system must use virtual SCSI disks only. Virtual disks have to be mapped to a physical storage that is actually located outside the source and the destination systems. No internal disks and no dedicated storage adapters are allowed.

If you want to migrate a running logical partition, the partition must use virtual Ethernet adapters and virtual SCSI disks provided by the Virtual I/O Server partition, and must not be assigned any physical adapters.

► Virtual I/O Server partition

There must be at least one Virtual I/O Server partition installed and activated on both the source and the destination systems.

Virtual SCSI adapter requirements:

– On the source Virtual I/O Server partition, do not set the adapter as required and do not select **Any client partition can connect** when you create a virtual SCSI adapter. The virtual SCSI adapter must be solely accessible by the client adapter of the mobile partition.

– On the destination Virtual I/O Server partition, do not create any virtual SCSI adapters for the mobile partition. These are created automatically by the migration function.

► Network connection

The mobile partition and the Virtual I/O Server partitions on the source and destination systems must be reachable from the HMC.

For a migration of running partitions (active migration), both the source and destination Virtual I/O Server partitions must be able to communicate with each other to transfer the mobile partition state. We suggest you use a dedicated network with 1 Gbps bandwidth, or more.

► Shared disks

One or more shared disks must be connected to the source and destination Virtual I/O Server partitions. At least one physical volume that is mapped by the Virtual I/O Server to a LUN on external SAN storage must be attached to the mobile partition.

The *reserve_policy* attribute of all the physical volumes that belong to the mobile partition must be set as no_reserve on the Virtual I/O Server partitions.

You can change this attribute by using **chdev** command on the Virtual I/O Server partition as follows:

```
$ chdev -dev hdiskX -attr reserve_policy=no_reserve
```

► Power supply

The destination system must be running on a regular power source. If the destination system is running on a battery power, return the system to its regular power source before migrating a partition.

4.1.1 Inactive partition migration

An inactive partition migration moves a powered-off partition from the source to the destination system together with its partition profile.

If a mobile partition has dedicated I/O adapters, it can only participate in the inactive partition migration. However, even in that case, the dedicated adapters are automatically removed from the partition profile so that the partition will boot with only virtual I/O resources after migration. If you need to use dedicated I/O adapters on the mobile partition after the migration, update the mobile partition's profile before booting, add adapters to the mobile partition by using dynamic LPAR operations, or make available the desired resources using other means.

> **Note:** It is a good practice to record the existing configuration at this point because the profile will be changed during a migration. This record can be used on the destination system to reconfigure any dedicated adapters.

4.1.2 Active partition migration

An active partition migration moves a running logical partition, including its operating system and applications, from the source to the destination system without disrupting the services of that partition.

The following requirements must be met for the active partition migration in addition to the minimal requirements on page 84:

► On both the source and the destination Virtual I/O Server partitions, one mover service partition (MSP) is enabled and one Virtual Asynchronous Services Interface (VASI) device is configured.

- No physical or dedicated I/O adapters are assigned to the mobile partition.
- Resource Monitoring and Control (RMC) connection between the mobile partition and the HMC is up.

Note: Any Virtual TTY sessions will be disconnected during the migration, and can be reestablished on the destination system by the user after migration.

4.2 New attributes and new virtual device of VIOS

For the Live Partition Mobility function, two new attributes and one virtual device are added for the Virtual I/O Server. Refer to 2.1, "Live Partition Mobility components" on page 18 for more detailed information.

4.2.1 Mover Service Partition

There must be at least one MSP on each of the source and the destination systems for a mobile partition to participate in an active partition migration. MSP is not required for inactive partition migration.

4.2.2 Virtual Asynchronous Services Interface device

The VASI device must be configured on both the source and the destination Virtual I/O Servers for the mobile partition to participate in an active partition migration. The VASI device is not required for inactive partition migration.

4.2.3 Time reference

Time reference is a new attribute of partitions that includes Virtual I/O Server partitions.

Synchronizing the time of day clocks for the source and destination Virtual I/O Server partitions is optional for both active and inactive partition migration. However, it is a recommended step for active partition migration. If you choose not to complete this step, the source and destination systems will synchronize the clocks while the mobile partition is moving from the source system to the destination system.

4.3 Preparing for an active partition migration

This section shows how to enable the Mover Service Partition (MSP) and define the Virtual Asynchronous Services Interface (VASI) device in turn. The MSP and the VASI device are required for the active partition migration only.

4.3.1 Enabling the Mover Service Partition

You can set the MSP attribute at the time you create a Virtual I/O Server partition, or dynamically for a running Virtual I/O Server.

In order to set the MSP attribute during the creation of a Virtual I/O Server partition, perform the following steps:

1. In the navigation pane, expand **Systems Management** → **Servers**, and select the system on which you want to create a new Virtual I/O Server partition.

2. In the Tasks pane (the lower right of the Hardware Management Console Workplace), expand **Configuration** → **Create Logical Partition**, and select **VIO Server**, as shown in Figure 4-2, to start the Create Lpar Wizard.
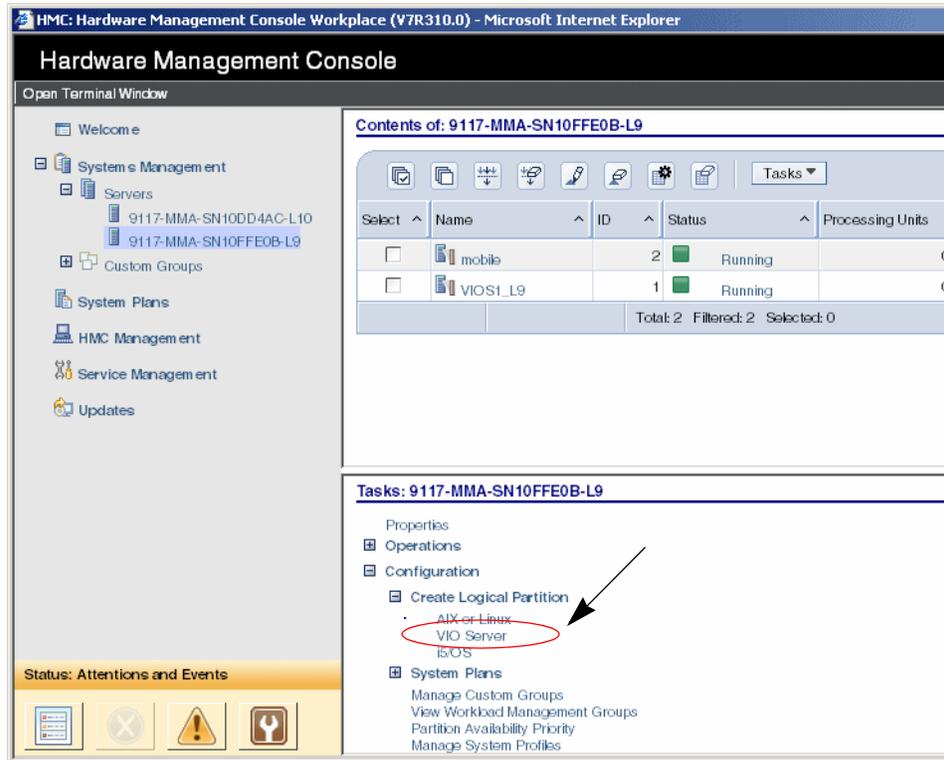
Figure 4-2   Hardware Management Console Workplace

3. Enter the partition name, change the ID if you like, and check the **Mover service partition** on the Create Lpar Wizard window (Figure 4-3).
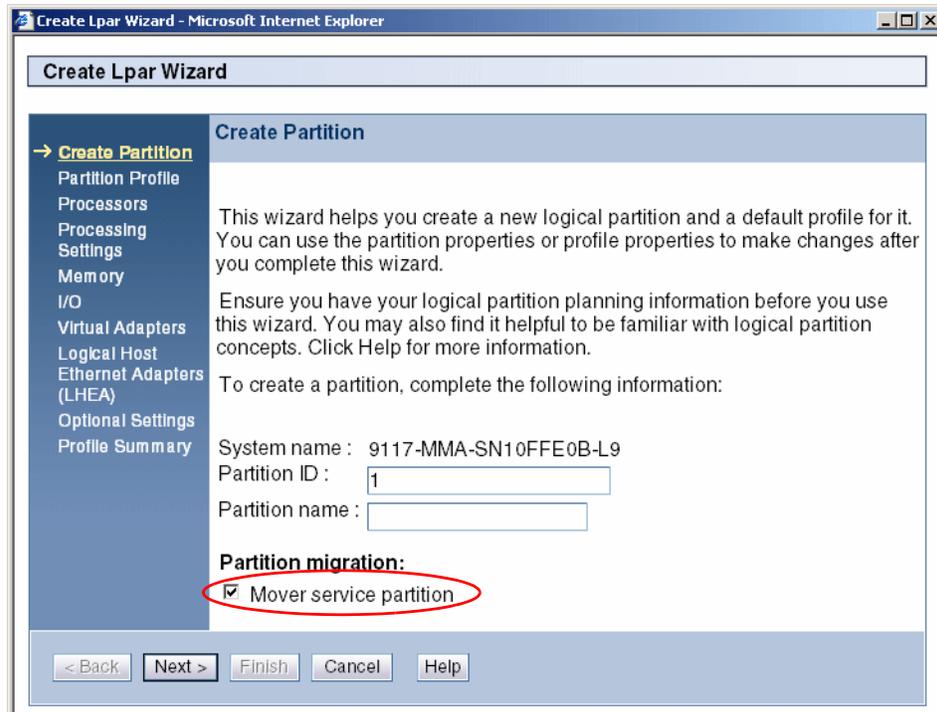
Figure 4-3   Create Lpar Wizard window

4. The mover service will be activated with the partition. Proceed with the steps of the Virtual I/O Server partition creation.

You can also set the MSP attribute dynamically for an existing Virtual I/O Server partition while the partition is in the **Running** state.

1. In the navigation pane, expand **Systems Management** → **Servers**, and select the desired system.

2. In the Contents pane (the top right of the Hardware Management Console Workplace), select the Virtual I/O Server for which you want to enable the MSP attribute.

3. Click **view popup menu** and select **Configuration** → **Properties**, as shown in Figure 4-4.
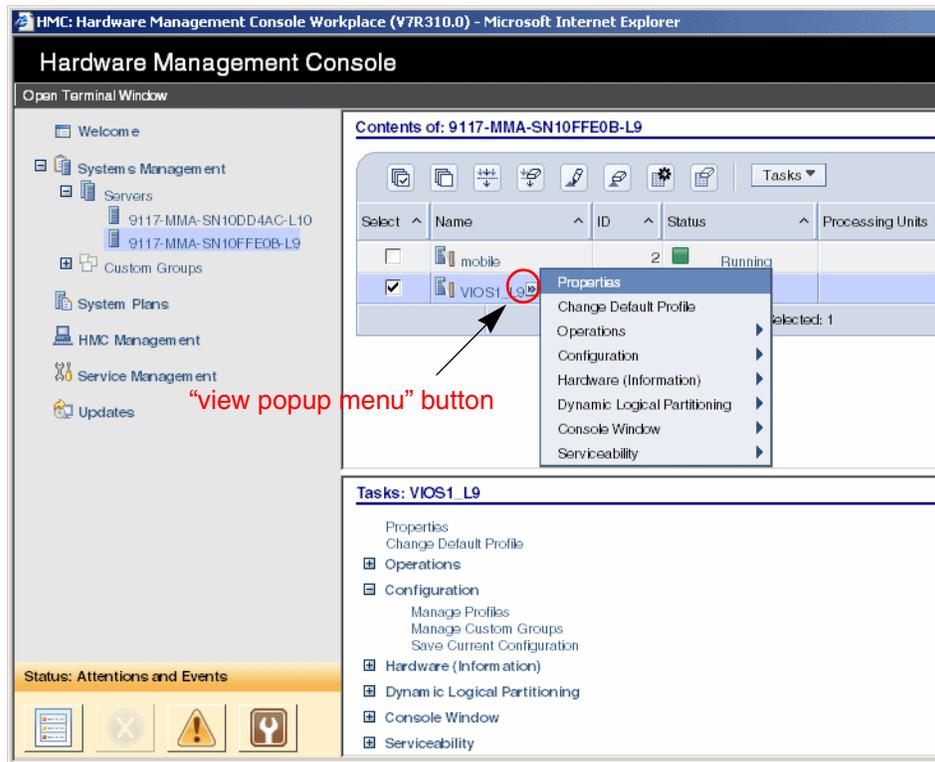
Figure 4-4   Changing the Virtual I/O Server partition property

4. Check the **Mover service partition** box on the **General** tab in the Partition Properties window, and press **OK** (Figure 4-5).
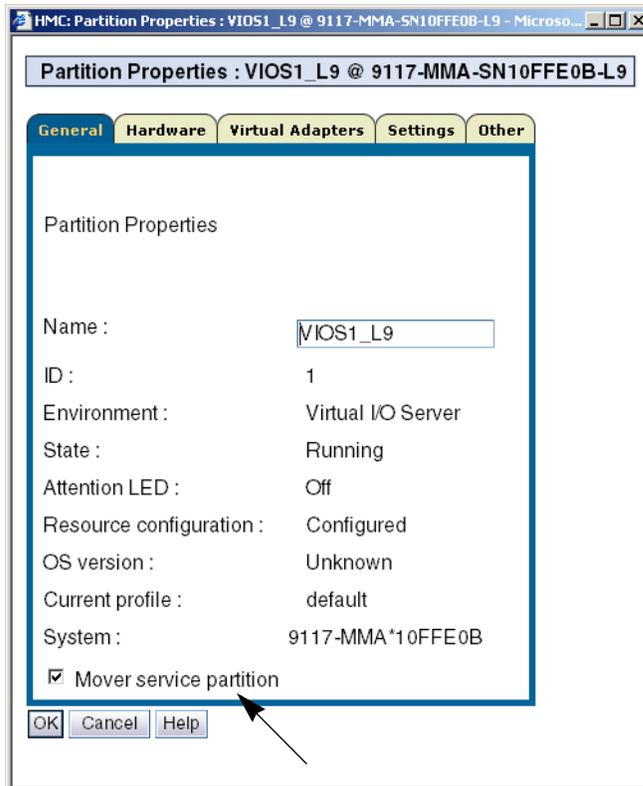
Figure 4-5   Enabling the Mover service partition attribute

4.3.2 Configuring a Virtual Asynchronous Services Interface device

You can define a VASI device at the time of creating a Virtual I/O Server partition, or dynamically just like virtual Ethernet and virtual SCSI adapters.

> **Important:** As of the November 2007 release, configuring the VASI device is no longer required. The VASI devices is automatically configured by the HMC.

In order to create the VASI device during the creation of a Virtual I/O Server partition, perform the following steps:

1. In the Create Lpar Wizard, proceed through the Virtual I/O Server configuration steps up to Virtual Adapters.

   Select **Virtual VASI** and press the **New**... button, as shown in Figure 4-6. Later HMC levels may have an updated Wizard.
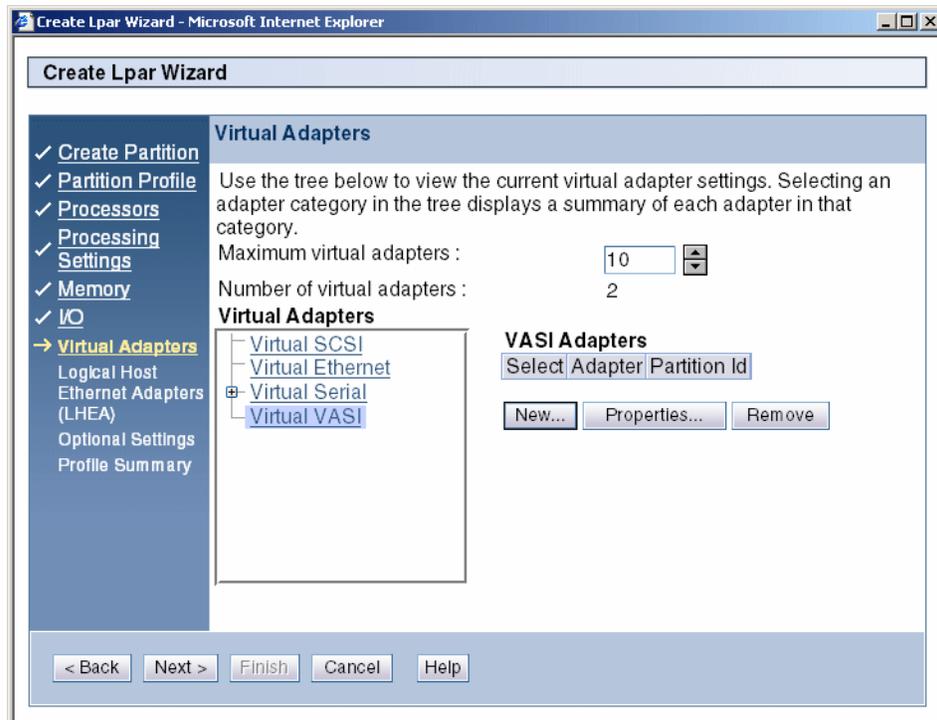
Figure 4-6   Creating a new VASI device

2. Change the number of **Maximum virtual adapters** if necessary, check the **Required adapter** box, and then click **Submit** (Figure 4-7).
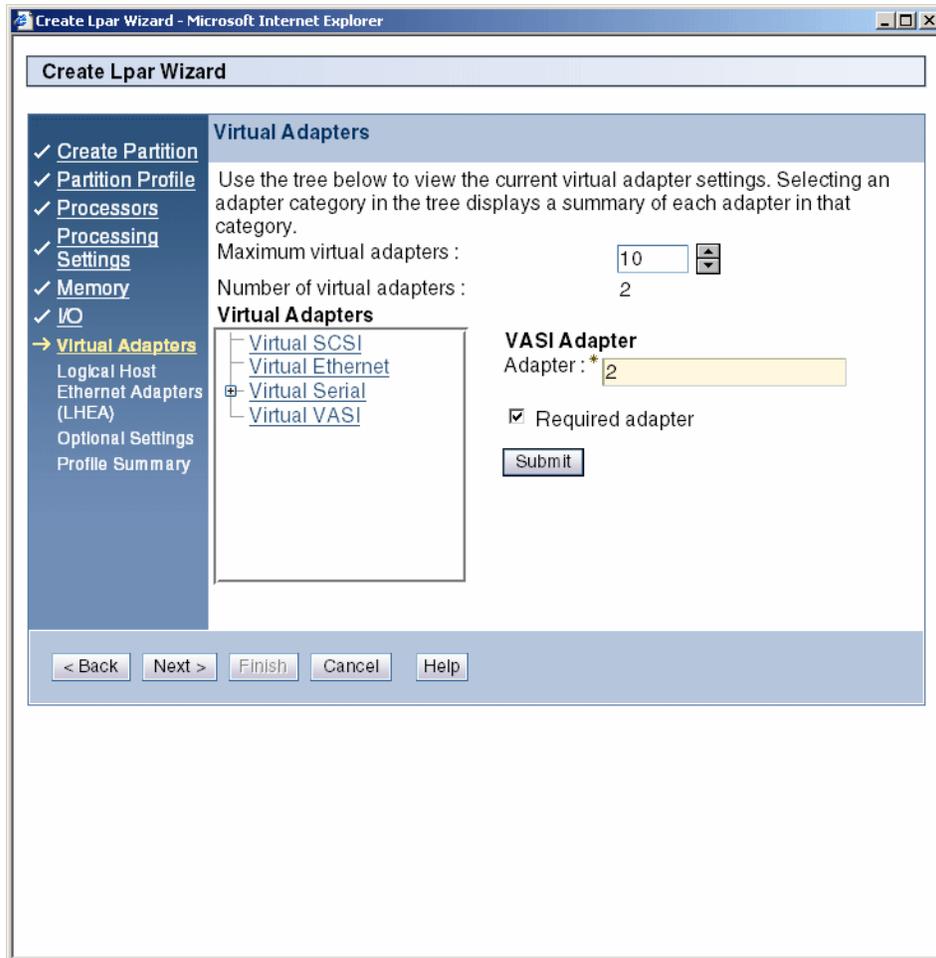

Figure 4-7   Creating the VASI device

3. Proceed and complete the remaining Virtual I/O Server configuration steps.

4. Install the Virtual I/O Server software to the new Virtual I/O Server partition.

You can also create a VASI device on the Virtual I/O Server partition while the partition is in the Running state by using the dynamic LPAR operation, or by modifying the partition profile and reactivating the partition.

On the Virtual I/O Server partition where a VASI device is configured, you can see a new vasi0 virtual device, as shown highlighted in Example 4-1.

Example 4-1   List of virtual devices

```
$ lsdev -virtual
name       status      description
ent1       Available  Virtual I/O Ethernet Adapter (l-lan)
ent2       Available  Virtual I/O Ethernet Adapter (l-lan)
ent3       Available  Virtual I/O Ethernet Adapter (l-lan)
vasi0      Available  Virtual Asynchronous Services Interface (VASI)
vhost0     Available  Virtual SCSI Server Adapter
vhost1     Available  Virtual SCSI Server Adapter
vsa0       Available  LPAR Virtual Serial Adapter
ent4       Available  Shared Ethernet Adapter
```

4.3.3 Enabling the Time reference

After creating a partition, you can enable the time reference attribute of the partition.

1. In the navigation pane, expand **Systems Management** → **Servers**, and select the system that has the partition for which you want to enable the time reference attribute.

2. In the Contents pane (the top right of the Hardware Management Console Workplace), select the partition. In this example, select the Virtual I/O Server partition.

3. Click **view popup menu** and select **Configuration** → **Properties**.

   The next steps are described in Figure 4-4 on page 91.

4. Select the **Settings** tab, select **Enabled** for the Time reference attribute, and press **OK**, as shown in Figure 4-8.
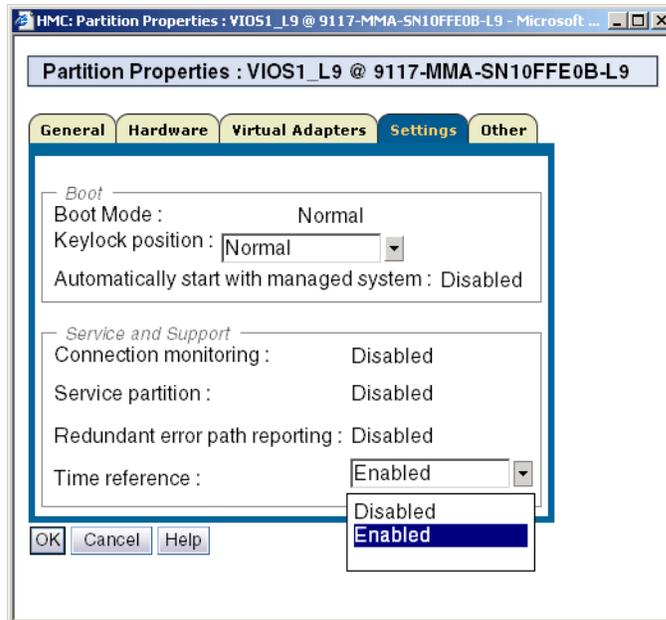

Figure 4-8   Enabling the Time reference attribute

4.4 Migrating a logical partition

This section shows detailed steps of how to migrate a logical partition that is called a *mobile partition* from the source to the destination system.

The main steps for the migration are the following:

► Perform the validation steps and eliminate errors.

► Inactive or active migration.

► Migrate the mobile partition.

The following sections show examples of mainly an active migration.

4.4.1 Performing the validation steps and eliminating errors

Before performing a migration, you should follow the validation steps. These steps are optional but recommended to eliminate errors. You can perform the validation steps by using the HMC GUI or CLI. In this section, we show the GUI steps. If you need more information about the CLI, refer to 5.6.1, "The migrlpar command" on page 139.

1. In the navigation pane, expand **Systems Management** → **Servers**, and select the source system.

2. In the contents pane (the top right of the Hardware Management Console Workplace), select the partition which you will migrate to the destination system.

3. Click **view popup menu** and select **Operations** → **Mobility** → **Validate**, as shown in Figure 4-9, to start the validation window.
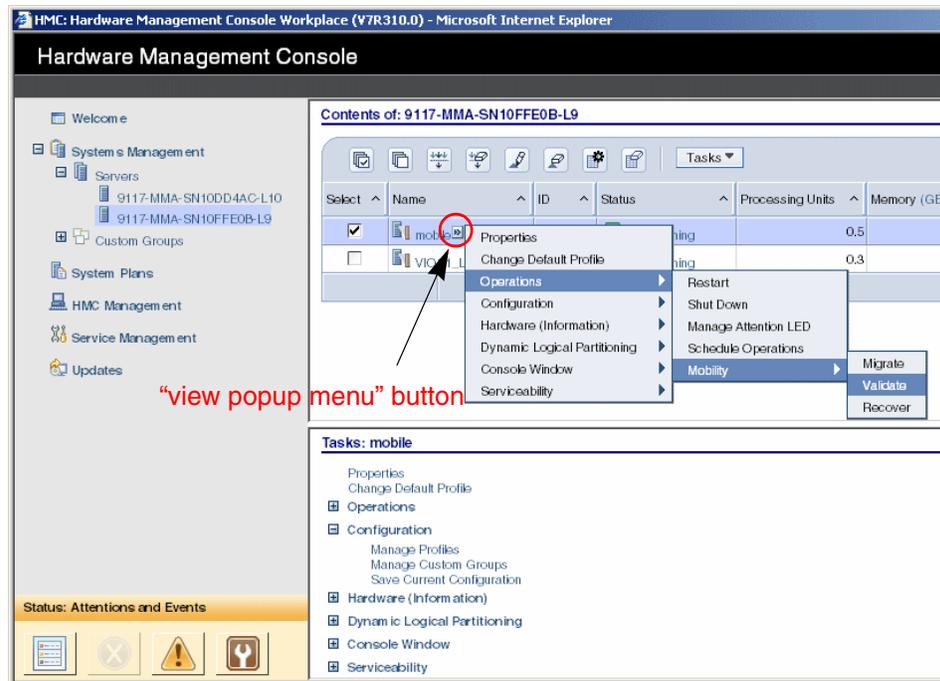
Figure 4-9   Validate menu on the HMC

4. Select the destination system, specify **Destination profile name** and **Wait time**, and then click the **Validate** button (Figure 4-10).

   If you are proceeding with this step when the mobile partition is in the Not Activated state, the Destination/Source mover service partition and wait time entries do not appear, because these are not required for the inactive partition migration.

Figure 4-10   Validation window

5. Check for errors or warnings in the Partition Validation Errors/Warnings window, and eliminate any errors.

   If there are any errors, check the messages in the window and the prerequisites for the migration. You cannot perform the migration steps with any errors. If there are warnings only, you may migrate the partition after the validation steps.

   For example, if you are proceeding with the validation steps on the mobile partition with physical adapters in the Running state (active migration), then you get the error shown in Figure 4-11.
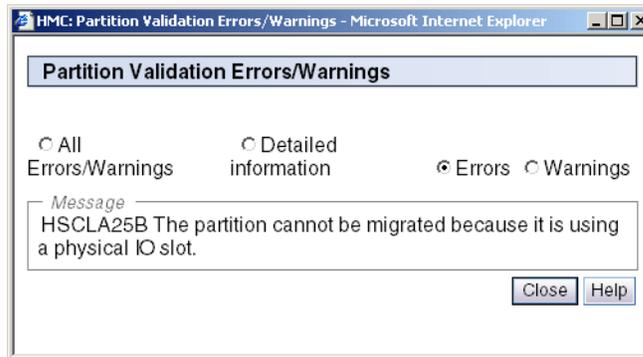

Figure 4-11   Partition Validation Errors

   If the mobile partition is in the Not Activated state, it reports a warning message, as shown in Figure 4-12.
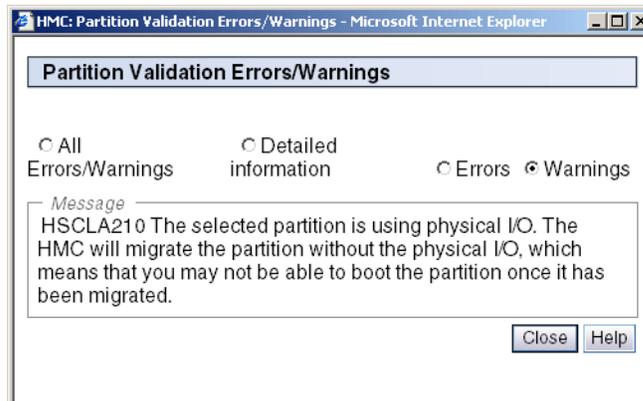

Figure 4-12   Partition Validation Warnings

6. After closing the **Partition Validation Errors/Warnings** window, a validation window as shown in Figure 4-13 appears again.

   If you have no errors in the previous step, you can also perform the migration at this point by clicking the **Migrate** button if you want.

Figure 4-13   Validation window after validation

4.4.2 Inactive or active migration

The migration type depends on the state of the mobile partition.

If you want to perform an inactive migration, the mobile partition should be powered off and in the Not Activated state.

If you want to perform an active migration, the mobile partition should be in the Running state, and no physical or dedicated I/O adapters must be assigned to it. For more details of the active partition migration requirements, refer to 4.1.2, "Active partition migration" on page 86.

4.4.3 Migrating a mobile partition

After the validation steps, migrate the mobile partition from the source to the destination system. You can perform the migration steps by using the HMC GUI or CLI. For more information about the CLI, refer to 5.6.1, "The migrlpar command" on page 139.

In this scenario, we are going to migrate a partition named *mobile* from the source system (9117-MMA-SN10FFE0B-L10) to the destination system (9117-MMA-SN10DD4AC-L10).

1. In the navigation pane, expand **Systems Management** → **Servers**, and select the source system.

   At this point, you can see the mobile partition is on the source system in Figure 4-14.
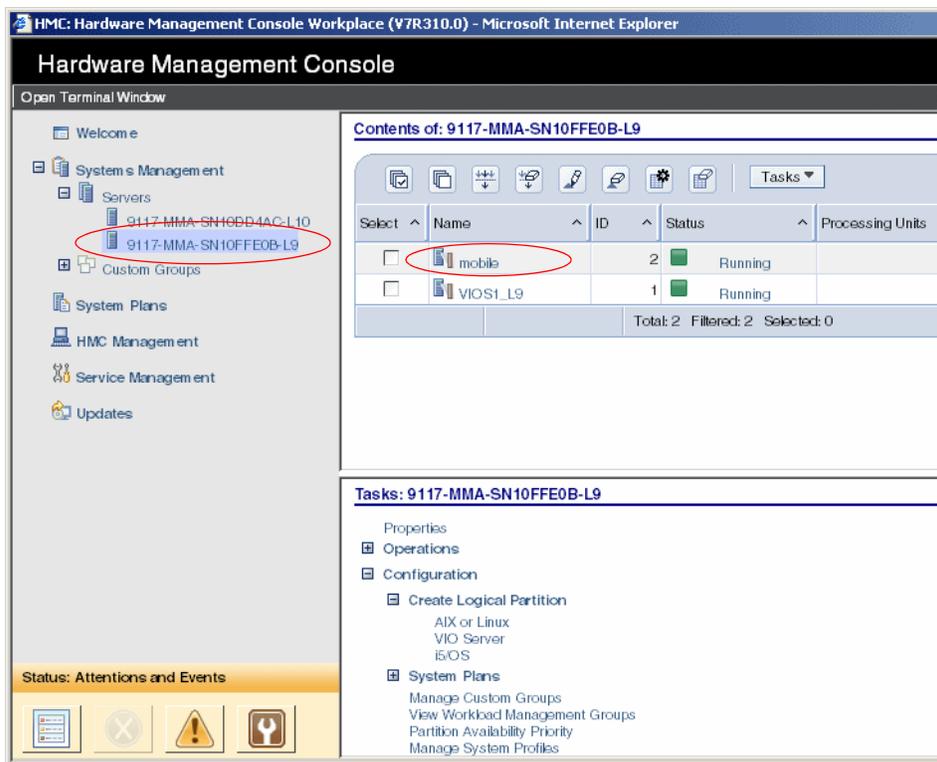

Figure 4-14   System environment before migrating

2. In the contents pane (the top right of the Hardware Management Console Workplace), select the partition that you will migrate to the destination system, that is, the mobile partition.

3. Click **view popup menu** and select **Operations** → **Mobility** → **Migrate**, as shown in Figure 4-15, to start the Partition Migration wizard.
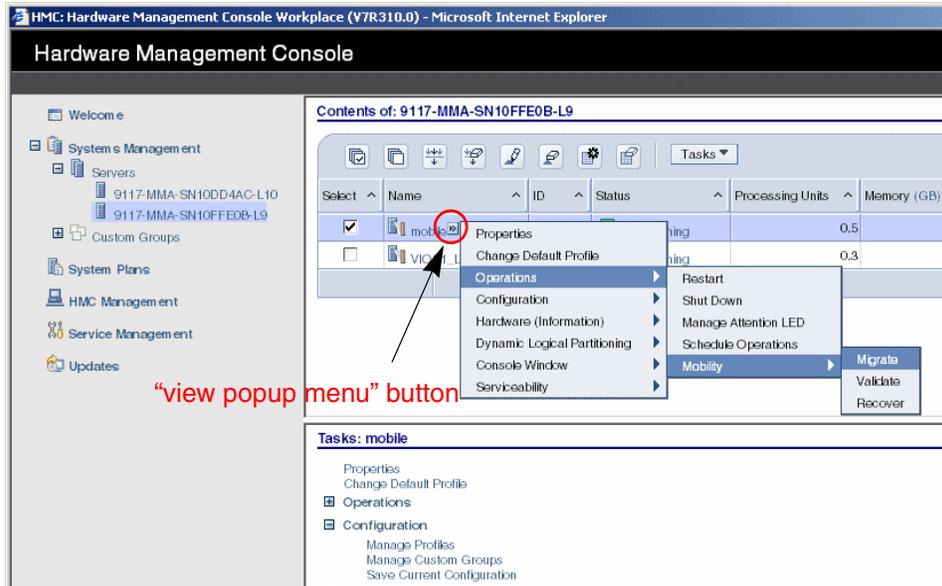

Figure 4-15   Migrate menu on the HMC

4. Check the Migration Information of the mobile partition in the Partition Migration wizard.

   If the mobile partition is powered off, Migration Type is Inactive. On the other hand, if the partition is in the Running state, it is Active (Figure 4-16).
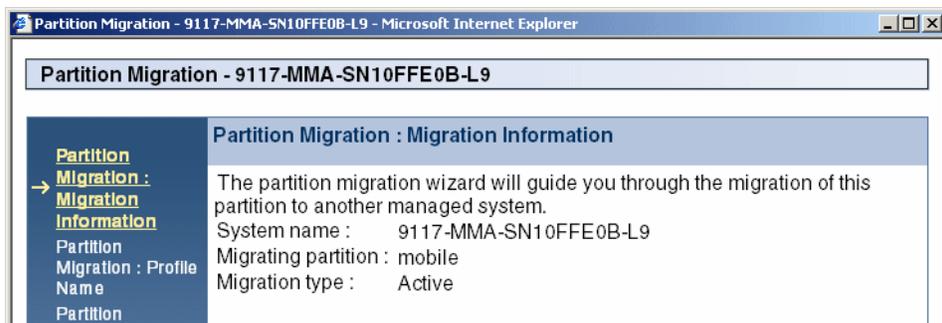

Figure 4-16   Migration Information

5. You can specify the New destination profile name in the Profile Name window (Figure 4-17).

   If you leave the name blank or do not specify a unique profile name, the profile on the destination system will be overwritten.
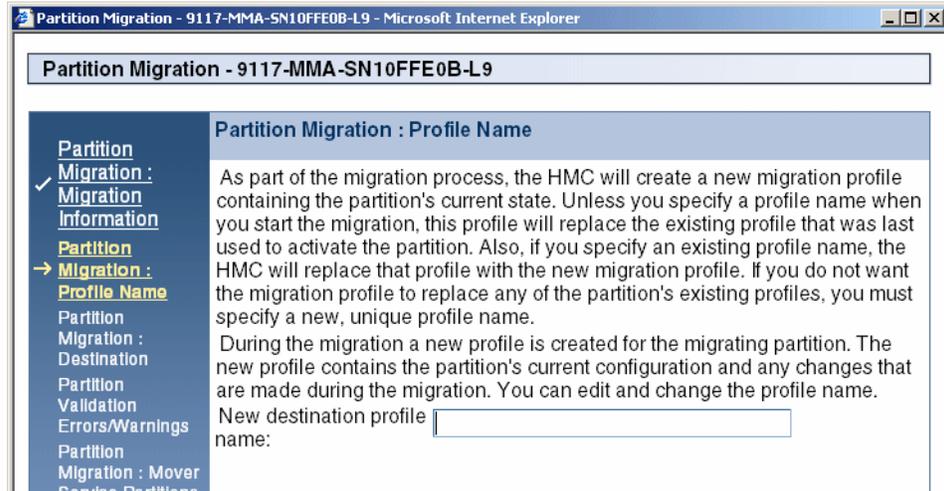

Figure 4-17   Specifying the profile name on the destination system

6. Select the destination system and click **Next** (Figure 4-18).
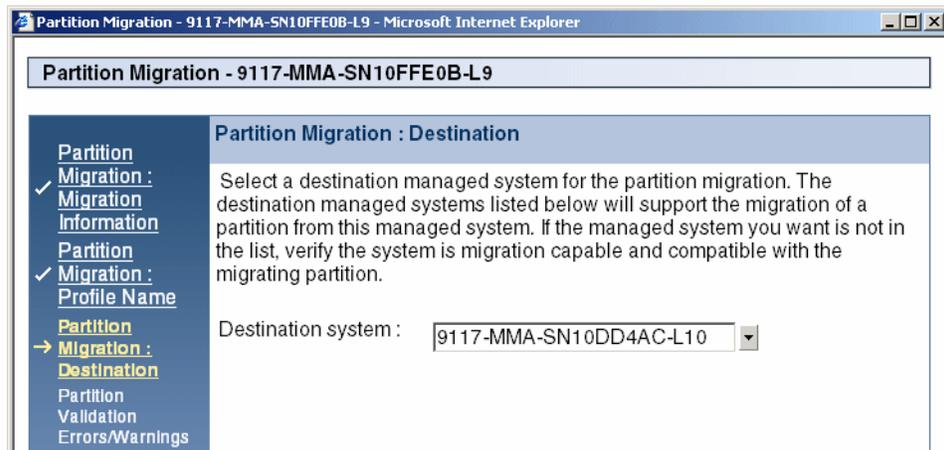
Figure 4-18   Selecting the destination system

   After this step, the HMC validates the partition migration environment.

7. Check errors or warnings in the Partition Validation Errors/Warnings window, and eliminate any errors. If there are any errors, you cannot proceed to the next step. You may proceed to the next step if it shows warnings only.

For example, if you do not specify a new destination profile name in a previous step (step 5 on page 103), the warning message shown in Figure 4-19 appears.
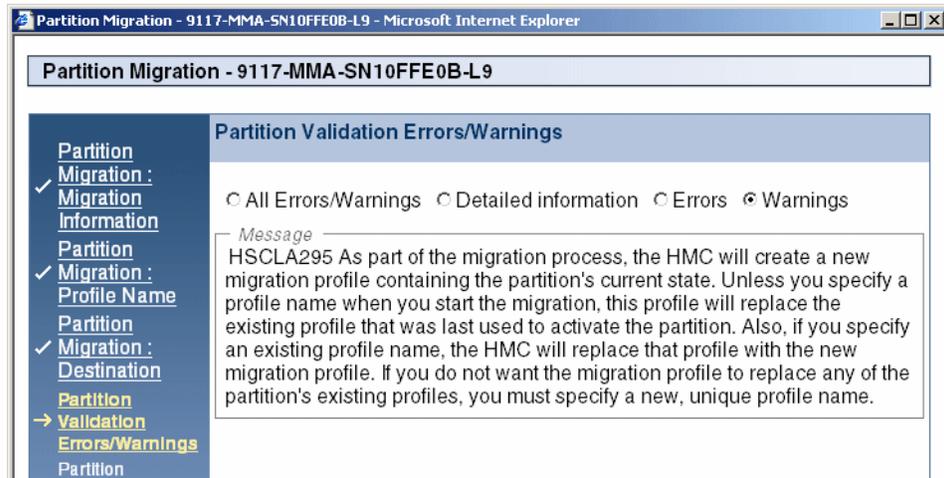
Figure 4-19   Sample of Partition Validation Errors/Warnings

8. Select the source and the destination **Mover Service Partitions** to be used for the migration (Figure 4-20 on page 105).
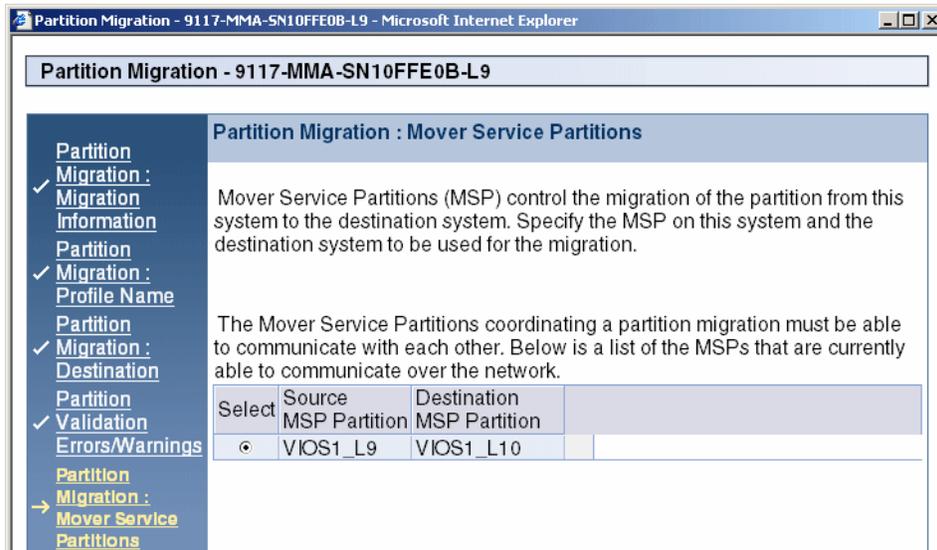
Figure 4-20   Selecting Mover Service Partitions

In this basic scenario one Virtual I/O Server partition is configured on each system, so this wizard window shows one candidate only. If you have more than one VIOS partitions on the source or on the destination system, you can select which mover server partitions to use.

If you are performing the inactive migration, this step is skipped (go to step 9 on page 106).

9. Select the VLAN configuration (Figure 4-21).
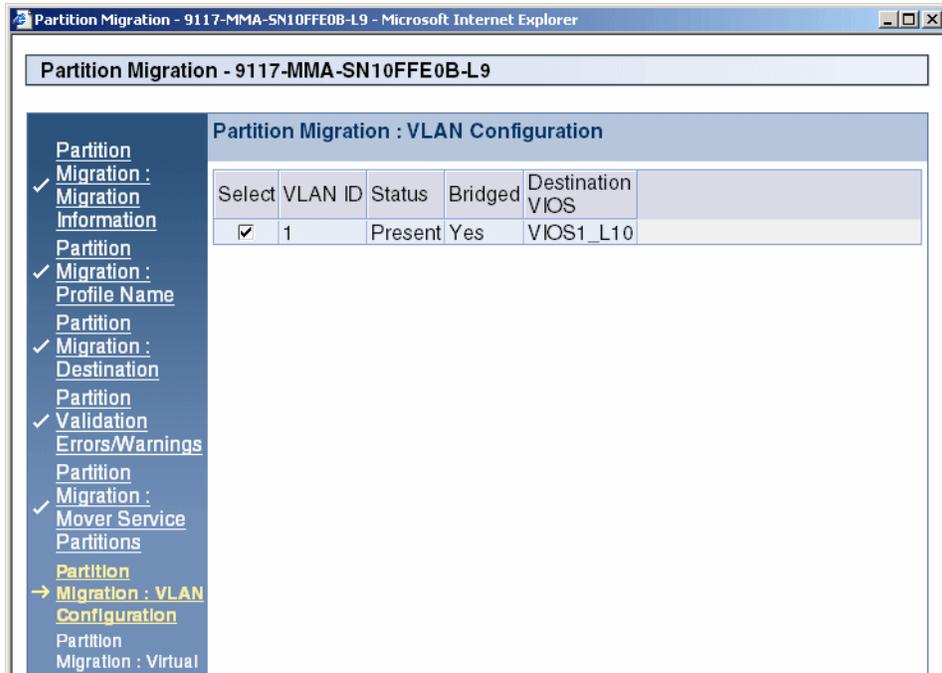
Figure 4-21   Selecting the VLAN configuration

Note: This step may not be available on some HMC maintenance levels.

10. Select the virtual SCSI adapter assignment (Figure 4-22).

    In this case, one Virtual I/O Server partition is configured on each system, so this wizard window shows one candidate only. If you have more than one Virtual I/O Server partition on the destination system, you may choose which Virtual I/O Server to use at the destination.
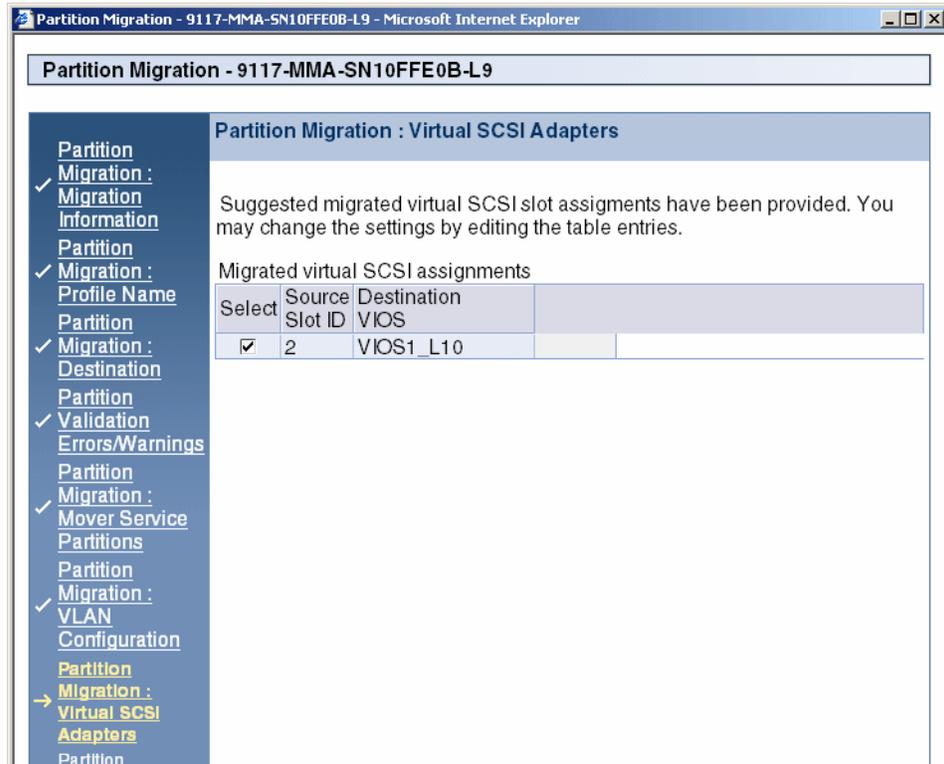
Figure 4-22   Selecting the virtual SCSI adapter

11. Specify the wait time in minutes (Figure 4-23 on page 108).

    The wait time value is passed to the commands that are invoked on the HMC and perform migration-related operations on the relevant partitions using the Remote Monitoring and Control (RMC).

    For example, the command syntax of **drmgr**, which can be used to install and configure dynamic logical partitioning (DLPAR) scripts, is:

```
drmgr {-i script_name [-w minutes] [-f] | -u script_name} [-D
hostname]
```

The wait time value is used for the argument for the -w option. If you specify 5 minutes as the wait time, the `drmgr` command is executed with -w 5.
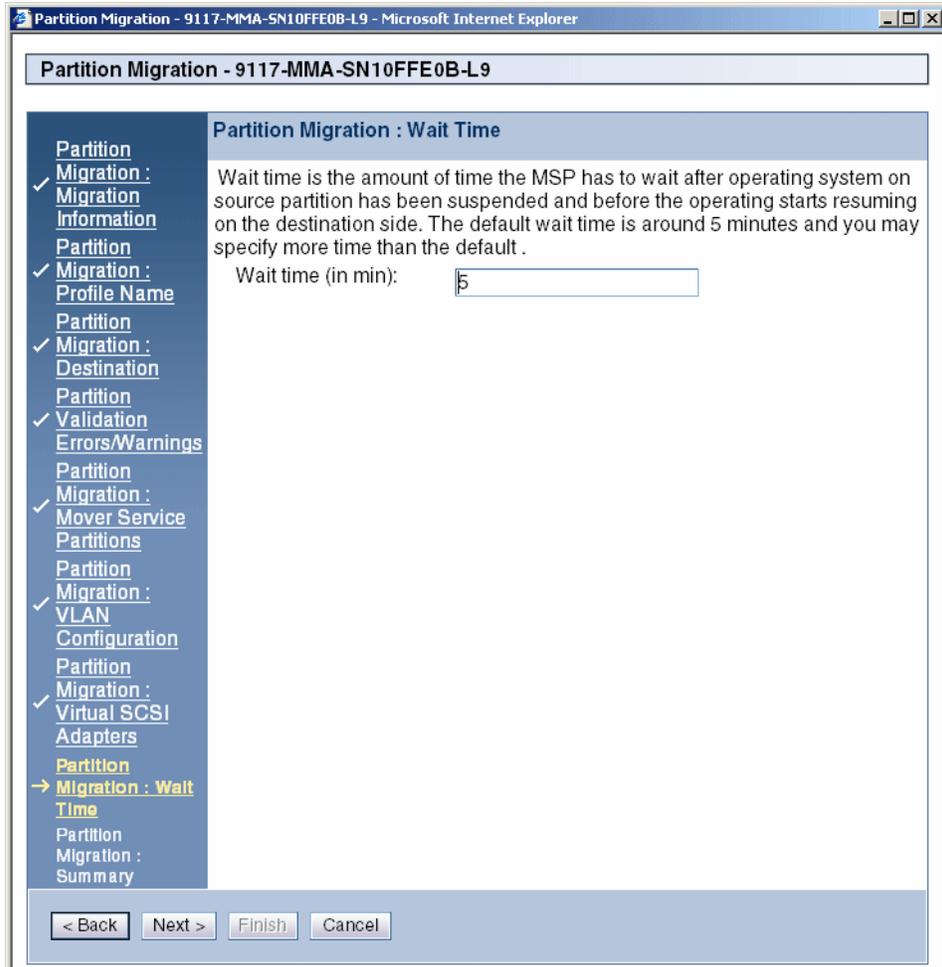
Figure 4-23   Specifying wait time

12. Check the settings that you have specified for this migration on the **Summary** window, and then press **Finish** to begin the migration (Figure 4-24).
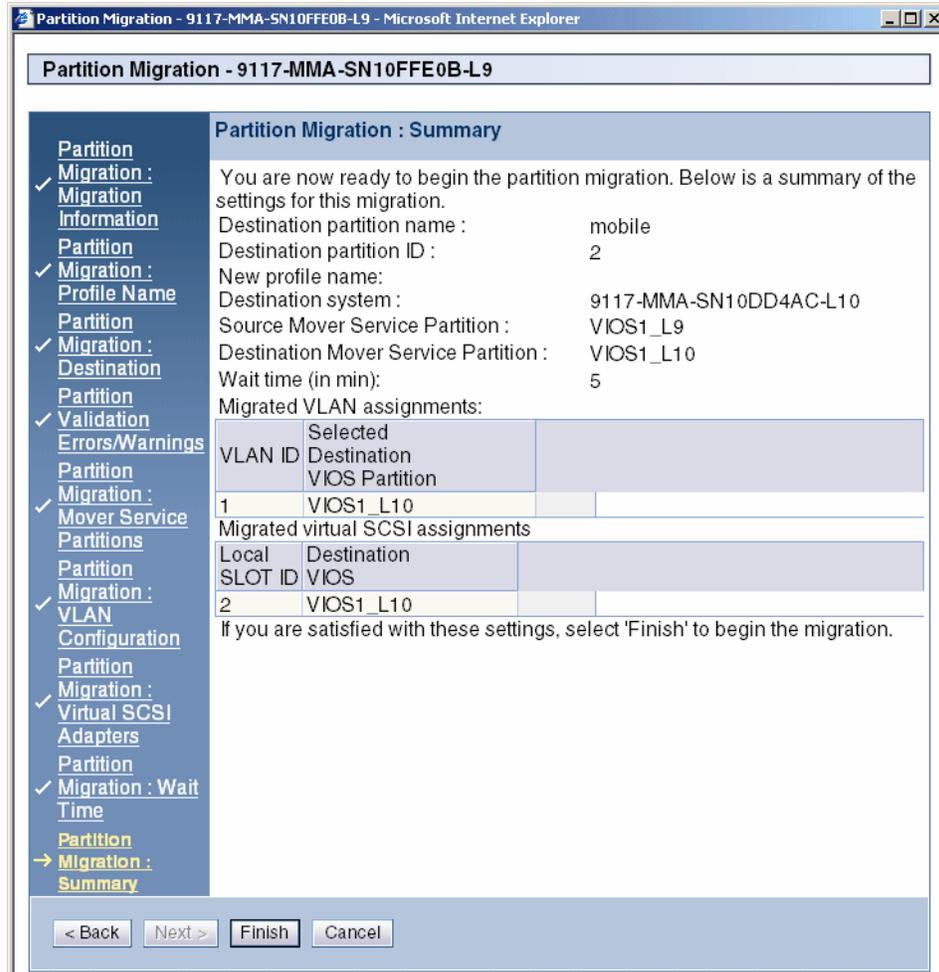
Figure 4-24   Partition Migration Summary window

13. You can see the Migration status and Progress (%) on the Partition Migration Status window (Figure 4-25).
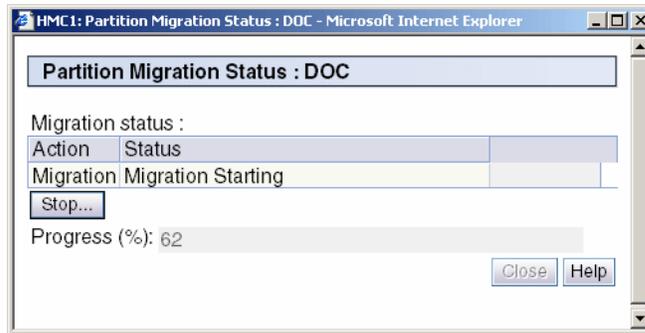
Figure 4-25   Partition Migration Status window

14. Check that the mobile partition is on the destination system.

You can see the mobile partition is on the destination system, as shown in Figure 4-26.
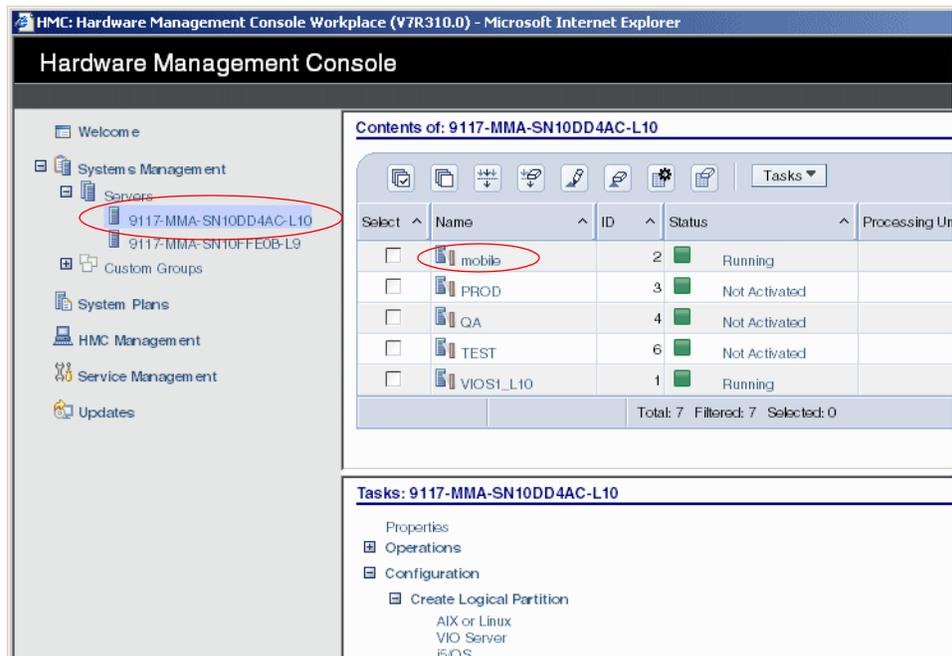
Figure 4-26   Migrated partition

15. If you keep a record of the virtual I/O configuration of the partitions, check the migrating partition's configuration in the destination system. While the migrating partitions keeps the same slot numbers as on the source systems, the server virtual adapter slot numbers may be different between the source and destination Virtual I/O Servers. Also, the virtual target device name may change during migration.

5

Advanced topics

This section contains various advanced topics relating to Live Partition Mobility. This chapter assumes you are familiar with the information in the preceding chapters. In this chapter, the following subjects are covered:

5.1 Dual Virtual I/O Server

Multiple Virtual I/O Servers are often deployed in systems where there is a requirement for logical partitions to continue to use their virtual resources even during the maintenance of a Virtual I/O Server.

This discussion relates to the common practice of using more than one Virtual I/O Servers allow for concurrent maintenance, and is not just limited to two servers. Also, Virtual I/O Servers may be created to offload the mover services to a dedicated partition.

Live Partition Mobility does not make any changes to the network setup on the source and destination systems. It only checks that all virtual networks used by the mobile partition have a corresponding shared Ethernet adapter on the destination system. Shared Ethernet failover may or may be not configured on either the source or the destination systems.

> **Important:** If you are planning to use shared Ethernet adapter failover, remember not to assign the Virtual I/O Server's IP address on the shared Ethernet adapter. Create another virtual Ethernet adapter and assign the IP address on it. Partition migration requires network connectivity through the RMC protocol to the Virtual I/O Server. The backup shared Ethernet adapter is always offline as well as its IP address, if any.

When multiple Virtual I/O Servers are involved, multiple virtual SCSI combinations may be possible, because access to the same SAN disk may be provided on the destination system by multiple Virtual I/O Servers. Live Partition Mobility automatically manages the virtual SCSI configuration if an administrator does not provide specific mappings.

The partition that is moving must keep the same number of virtual SCSI adapters after migration and each virtual disk must remain connected to the same adapter or adapter set. An adapter's slot number may change after migration, but the same device name is kept by the operating system for both adapters and disks.

A migration may fail validation checks and is not started if the moving partition adapter and disk configuration cannot be preserved on the destination system. In this case, you are required to modify the partition configuration before starting the migration.

> **Tip:** We suggest that you always perform a validation before performing a migration. The validation checks the configuration of the involved Virtual I/O Servers and shows you the configuration that will be applied. Use the validation menu on the GUI or the `lslparmigr` command described in 5.6, "The command-line interface (CLI)" on page 138.

In this section, we describe three different migration scenarios where the source and destination systems provide disk access either with one or two Virtual I/O Servers.

5.1.1 Dual Virtual I/O Server and client mirroring

Dual Virtual I/O Server and client mirroring may be used when you have two independent storage subsystems providing disk space with data mirrored across them. With this setup, the partition can keep running if one of the subsystems is taken offline.

If the destination system has two Virtual I/O Servers, one of them must be configured to access the disk space provided by the first storage subsystem while the other must access the second subsystem, as shown in Figure 5-1.
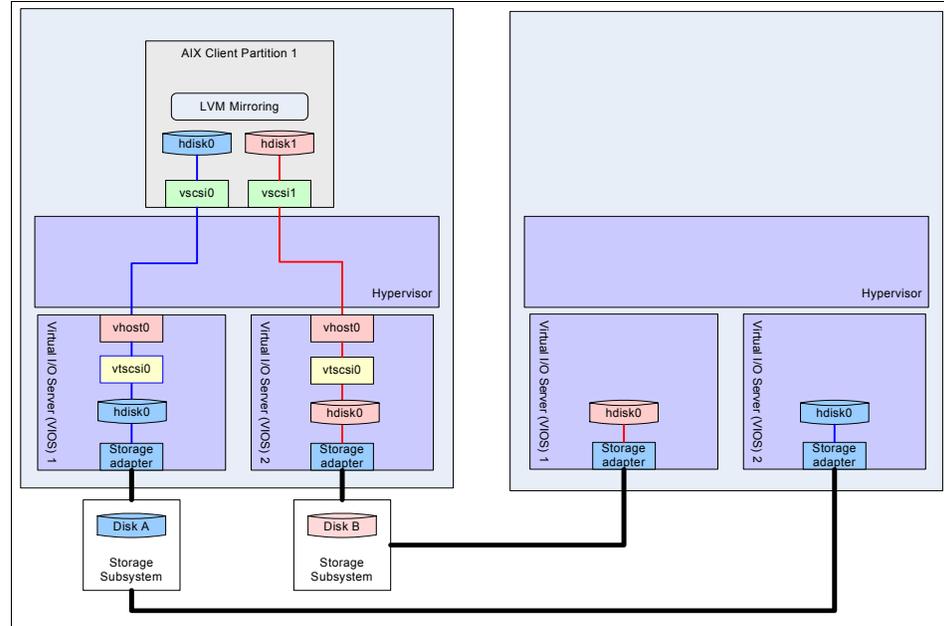
Figure 5-1   Dual VIOS and client mirroring to dual VIOS before migration

The migration process automatically detects which Virtual I/O Server has access to which storage and configures the virtual devices to keep the same disk access topology.

When migration is complete, the logical partition has the same disk configuration it had on previous system, still using two Virtual I/O Servers, as shown in Figure 5-2.
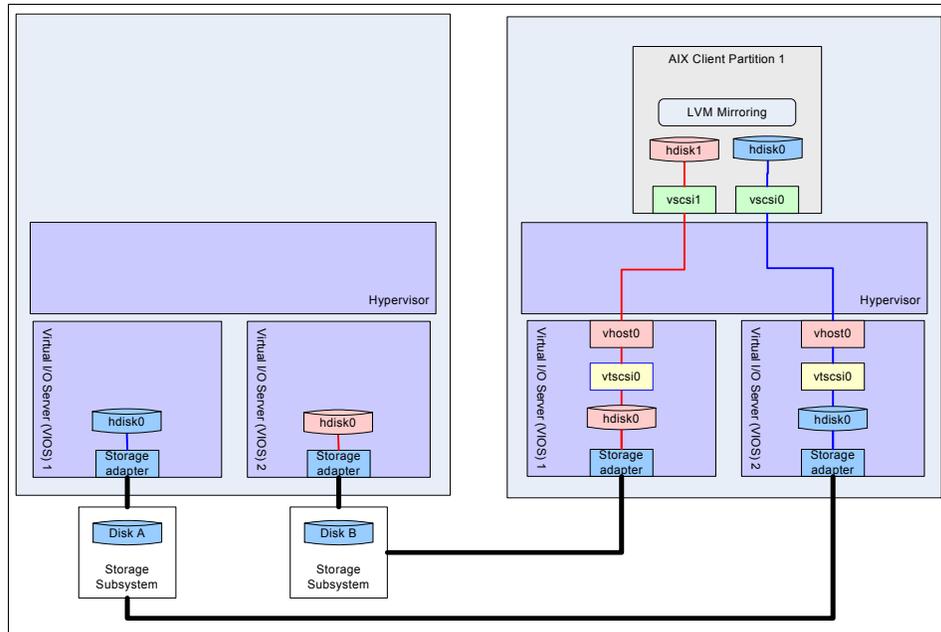
Figure 5-2   Dual VIOS and client mirroring to dual VIOS after migration

If the destination system has only one Virtual I/O Server, the migration is still possible and the same virtual SCSI setup is preserved at the client side. The destination Virtual I/O Server must have access to all disk spaces and the process creates two virtual SCSI adapters on the same Virtual I/O Server, as shown in Figure 5-3.
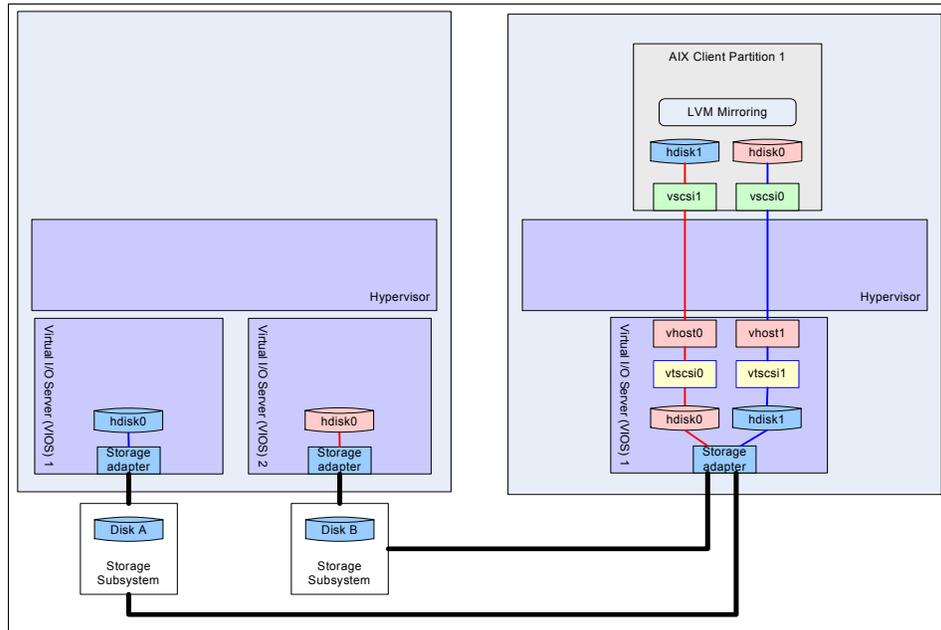
Figure 5-3   Dual VIOS and client mirroring to single VIOS after migration

5.1.2 Dual Virtual I/O Server and multipath I/O

With multipath I/O, the logical partition accesses the same disk data using two different paths, each provided by a separate Virtual I/O Server. One path is active and the other is standby.

The migration is possible only if the destination system is configured with two Virtual I/O Servers that can provide the same multipath setup. They both must have access to the shared disk data, as shown in Figure 5-4.
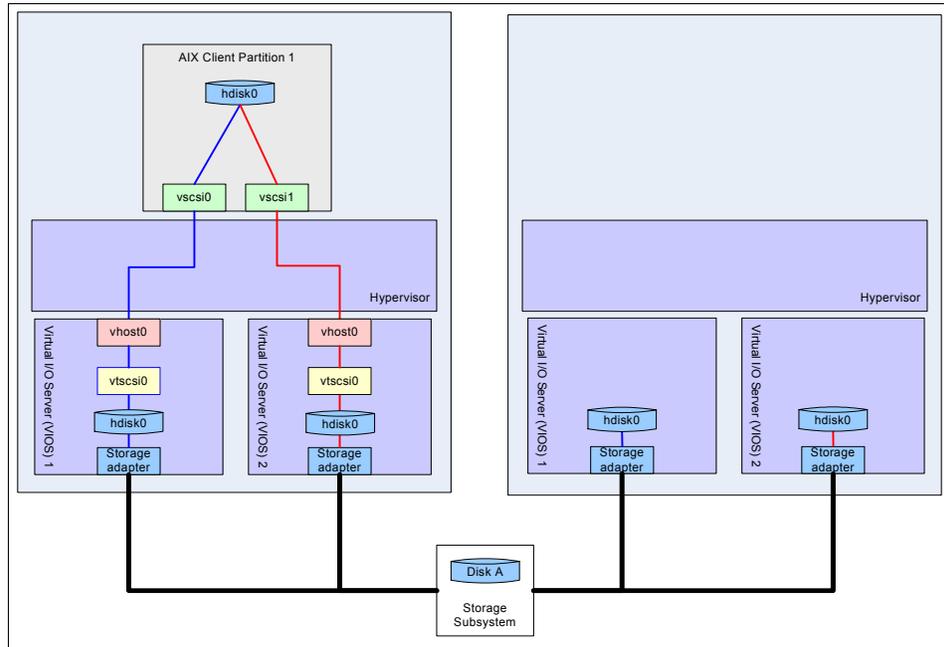
Figure 5-4   Dual VIOS and client multipath I/O to dual VIOS before migration

When migration is complete, on the destination system, the two Virtual I/O Servers are configured to provide the two paths to the data, as shown in Figure 5-5.
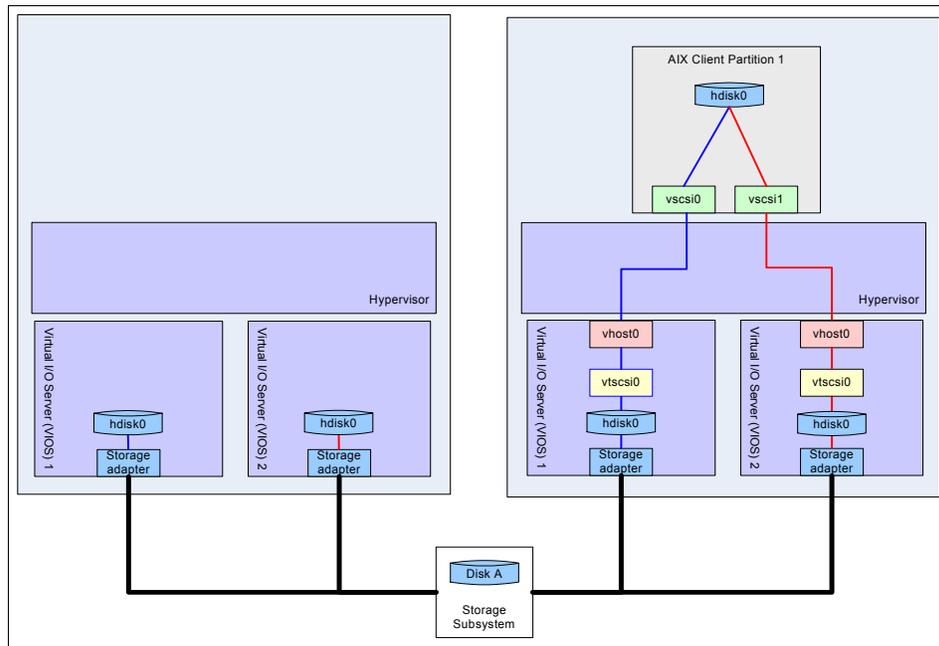
Figure 5-5   Dual VIOS and client multipath I/O to dual VIOS after migration

If the destination system is configured with only one Virtual I/O Server, the migration cannot be performed. The migration process would create two paths using the same Virtual I/O Server, but this setup is not allowed, because it is not recommended to have two virtual target devices that map the same backing device on different virtual SCSI server devices.

To migrate the partition, you must first remove one path from the source configuration before starting the migration. The removal can be performed without interfering with the running applications. The configuration becomes a simple single Virtual I/O Server migration.

5.1.3 Single to dual Virtual I/O Server

A logical partition that is using only one Virtual I/O Server for virtual disks may be migrated to a system where multiple Virtual I/O Servers are available.

Since the migration never changes a partition's configuration, only one Virtual I/O Server is used on the destination system.

If access to all disk data required by the partition is provided by only one Virtual I/O Server on the destination system, after migration the partition uses just that Virtual I/O Server. If no destination Virtual I/O Server provides all disk data, the migration cannot be performed.

When both destination Virtual I/O Servers have access to all the disk data, the migration can select either one or the other. When you start the migration, you have the option of choosing a specific Virtual I/O Server; otherwise, the HMC automatically makes a selection. The situation is shown in Figure 5-6.
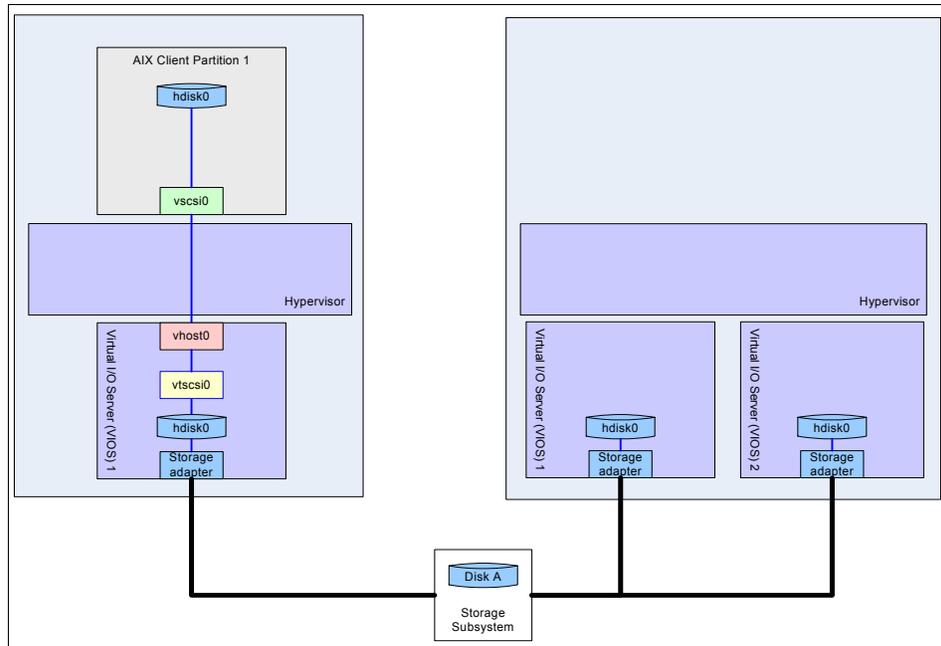
Figure 5-6   Single VIOS to dual VIOS before migration

When the migration is performed using the GUI on the HMC, a list of possible Virtual I/O Servers to pick from is provided. The command-line interface by default makes the automatic selection if no specific option is provided.

After migration, the configuration is similar to the one shown in Figure 5-7.
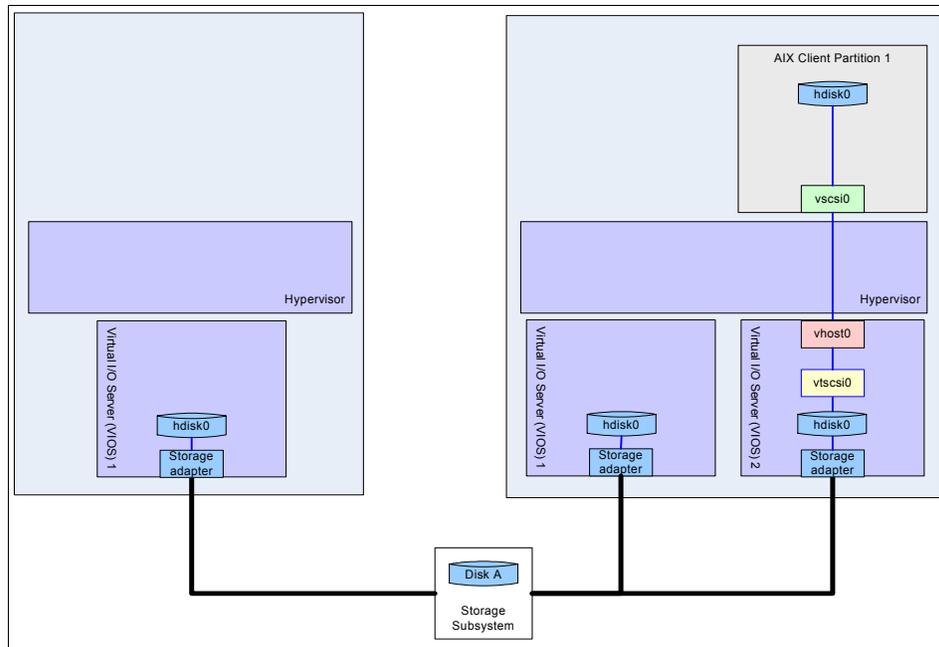
Figure 5-7   Single VIOS to dual VIOS after migration

5.2 Multiple concurrent migrations

The same system can handle multiple concurrent partition migrations, any mix of either inactive or active.

There are many scenarios where more than one migration may be started on the same system. For example:

► A review of the entire infrastructure detects that a different system location of some logical partition may improve global system usage and service quality.

► A system is planned to enter maintenance and must be shut down. Some of its partitions cannot be stopped or the planned maintenance time is too long to satisfy service level agreements.

The maximum number of concurrent migrations on a system can be identified using the `lslparmigr` command on the HMC with the following syntax:

```
lslparmigr -r sys -m <system>
```

Some practical considerations should be taken into account when planning for multiple migrations, especially when there is the need to evaluate the time required by the migration process.

For each mobile partition, you must use an HMC GUI wizard or an HMC command. While a migration is in progress, you can start another one. When the number of migrations to be executed grows, the setup time using the GUI may become long and you should consider using the command-line interface. The `migrlpar` command may be used in scripts to start multiple migrations in parallel.

An active migration requires more time to complete than an inactive migration because the system performs additional activities to keep applications running while the migration is in progress.

- ► The time required to complete an active migration depends on the size of the memory to be migrated and on the mobile partition's workload.
- ► The Virtual I/O Servers selected as mover service partitions are loaded by memory moves and network data transfer.
  - – High speed network transfers may become CPU intensive workloads.
  - – At most, four concurrent active migrations may be managed by the same mover service partition.

The active migration process has been designed to handle any partition memory size and it is capable of managing any memory workload. Applications can update memory with no restriction during migration and all memory changes are taken into account, so elapsed migration time may change with workload. While the algorithm is efficient, planning the migration during low activity periods may help to reduce migration time.

Virtual I/O Servers selected as mover service partitions are involved in partition's memory migration and must manage high network traffic. Network management can cause high CPU usage and usual performance considerations apply: use uncapped Virtual I/O Servers and add virtual processors if the load increases. Alternatively, create dedicated Virtual I/O Servers on the source and destination systems that provide the mover service function separating the service network traffic from the migration network traffic. You can combine or separate virtualization functions and mover service functions to suit your needs.

If multiple mover service partitions are available on either the source or destination systems, we suggest distributing the load among them. This can be done explicitly by selecting the mover service partitions either using the GUI, or the command-line interface. Each mover service partition can manage up to four concurrent active migrations and explicitly using multiple Virtual I/O Servers avoids queuing of requests.

5.3 Dual HMC considerations

The HMC is the center of system management of IBM System p servers and its unavailability does not affect service by any means. It is, however, possible to deploy two HMCs managing the same systems.

In a dual HMC configuration, both HMCs see the same system's status, have the same configuration rights, and can perform the same actions. In order to avoid concurrent operations on the same system, a locking mechanism is in place that allows the first configuration change to occur and the second one to fail with a message showing the identifier of the locking HMC.

Partition migration is a configuration change that involves two separate systems. Moreover, the migration process requires that no additional modifications occur on the involved objects until migration is completed.

The HMC that initiates a migration takes a lock on both managed systems and the lock is released when migration is completed. The other HMC can show the status of migration but cannot issue any additional configuration changes on the two systems. The lock can be manually broken, but this option should be considered carefully.

The locking mechanism must be taken into account when migration is planned. Most additional system management actions should be performed using the same HMC that is performing the migration. The other HMC can still be used for monitoring purposes, but not for configuration changes until migration is completed.

When multiple migrations are planned between two systems, multiple HMC commands are issued. The first migration task takes an HMC lock on both systems so the following migration must be issued on the same HMC. It is then required that only one HMC is used when multiple concurrent migrations are executed.

5.4 Multiple shared-processor pools

IBMs POWER6-based System p servers with AIX Version 6.1 support multiple shared-processor pools. The migration wizard presents you with a list of all defined shared-processor pools on the destination that have sufficient capacity to receive the migrating partition. You are asked to identify the target pool as shown in Figure 5-8 on page 124. The name and identifier of the shared-processor pool on the destination do not have to be the same as those on the source. If you use the command-line then the migration operation will fail if the arrival of the

migrating partition would cause the maximum processors in the chosen shared pool on the destination to be exceeded.
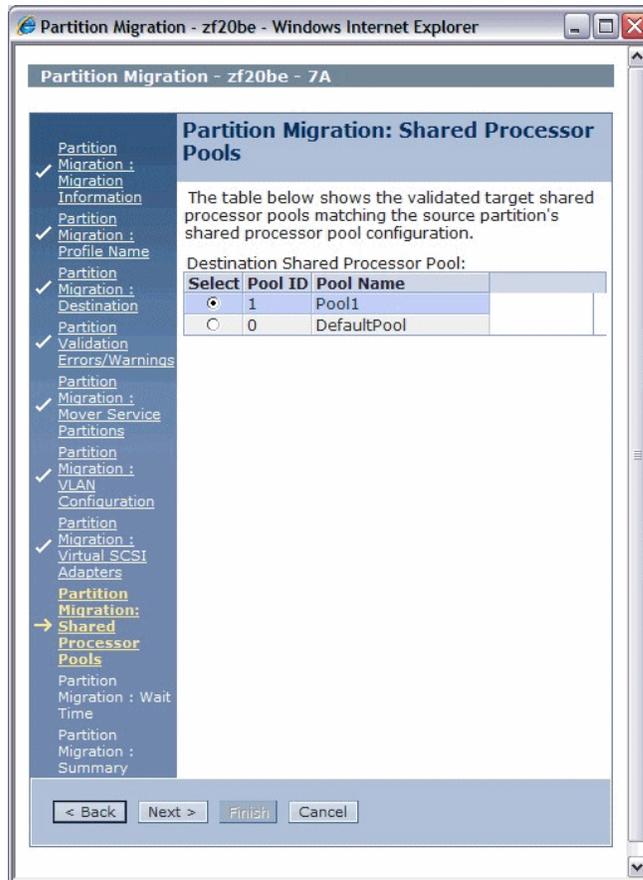
Figure 5-8   Partition Migration: Shared Processor Pools

5.5 Migrating a partition with physical resources

This section explains how to migrate a partition that is currently using physical resources.

5.5.1 Overview

Three types of adapters cannot be present in a partition when it is participating in an active migration: physical adapters, HEA, and non-default virtual serial adapters. A non-default virtual serial adapter is a virtual serial adapter other than the two automatically created virtual serial adapters in slots 0 and 1. If a partition has any such adapters, you must deconfigure these and you may need to switch from physical to virtual resources.

We assume for this scenario that you are beginning with a mobile partition that is using a single physical Ethernet adapter and a single physical SCSI adapter (see Figure 5-9).

Figure 5-9   The mobile partition is using physical resources

If the mobile partition has any adapters that cannot be migrated, then they must be removed from the mobile partition before it can participate in an active migration. If these adapters are marked as desired in the active profile, remove them using dynamic logical partitioning. If these adapters are marked as required in the active profile, activate the partition with a profile that does not have them marked as required.
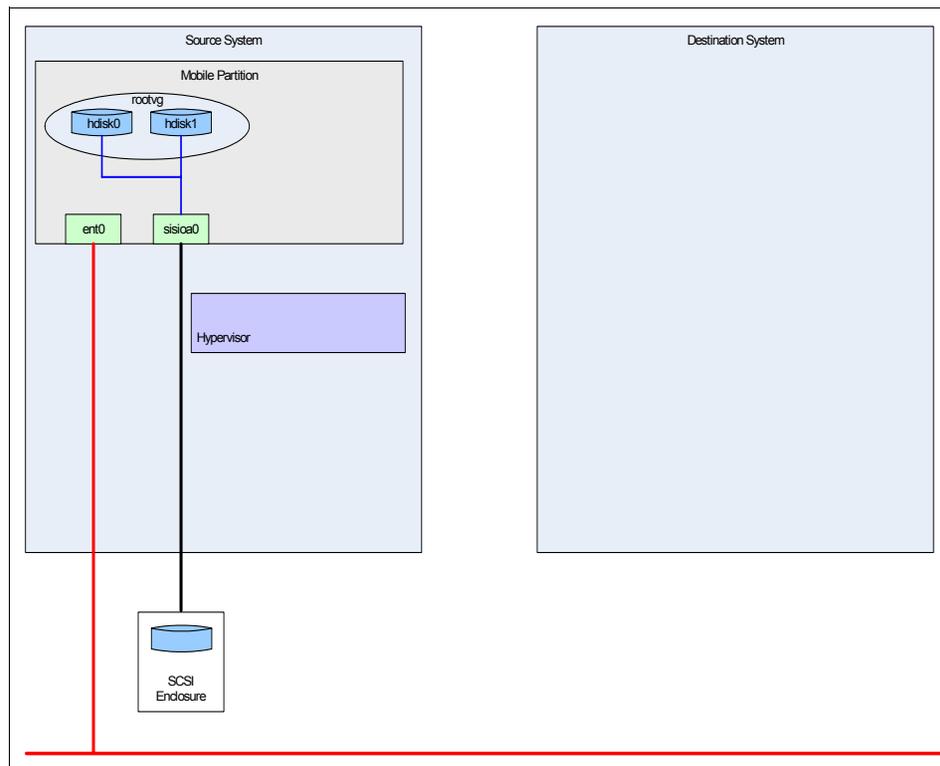
The process described in this section covers both the case where the mobile partition has no such required adapters, and the case where it does.

Before proceeding, verify that you meet the requirements for Live Partition Mobility, as outlined in Chapter 3, "Requirements and preparation" on page 41. Ignore the stricture against adapters that cannot be migrated, because this scenario describes how to deal with them.

5.5.2 Configure a Virtual I/O Server on the source system

Create and install a Virtual I/O Server partition on the source system. When creating and configuring the partition, refer to the following procedure. For detailed instructions for these steps, consult the basic configuration guide in 4.3, "Preparing for an active partition migration" on page 88.

► Configure a virtual SCSI server adapter.

> **Important:** Mark the virtual SCSI server adapter as desired (not required) in your Virtual I/O Server partition profile. This is necessary to allow the migration process to dynamically remove this adapter during a migration.
>
> When creating the virtual SCSI server adapter, use the Only selected client partition can connect option. For the Client partition field, specify the mobile partition. For the Client adapter field, specify an unused virtual slot on the mobile partition. Do not set the server adapter to accept connections from any partition. This allows the migration process to identify which server adapter is paired with which client partition.

► Attach and configure the remote storage using a storage area network.

– Create one LUN on your storage subsystem for each disk in use on your mobile partition. Ensure that these LUNs are at least as large as the disks on your mobile partition. Make these LUNs available as hdisks on the source Virtual I/O Server.

– On the source Virtual I/O Server, set the reserve_policy on the disks to no_reserve using the **chdev** command:

```
$ chdev -dev hdisk5 -attr reserve_policy=no_reserve
```

– Assign the hdisks as targets of the virtual SCSI server adapter that you created, using the `mkvdev` command. Do not create volume groups and logical volumes on the hdisks within the Virtual I/O Server.

► Configure shared Ethernet adapters for each physical network interface that is configured on the mobile partition.

► Create one VASI device for the Virtual I/O Server.

► Ensure the **Mover service partition** box is checked in the Virtual I/O Server partition properties.

Figure 5-10 shows the created and configured source Virtual I/O Server.
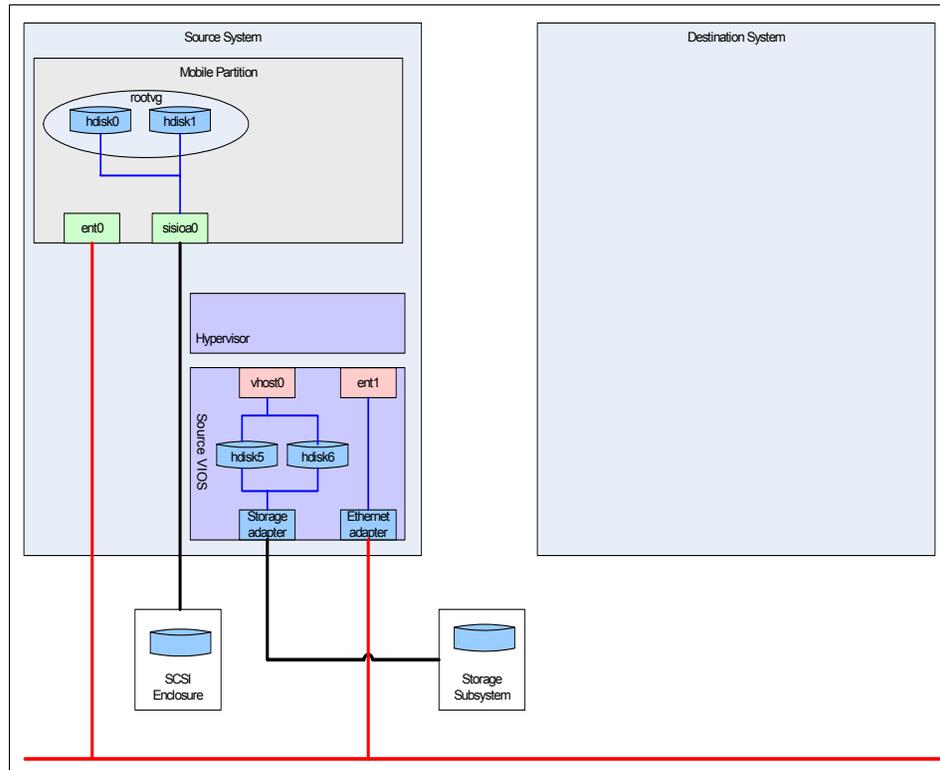

Figure 5-10   The source Virtual I/O Server is created and configured

5.5.3 Configure a Virtual I/O Server on the destination system

Create and install a Virtual I/O Server partition on the destination system. When creating and configuring the partition, refer to the following instructions. For detailed instructions for these steps, consult the basic configuration guide in 4.3, "Preparing for an active partition migration" on page 88.

> **Important:** Do not create any virtual SCSI server adapters for your mobile partition on the destination Virtual I/O Server. Do not map any shared hdisks on the destination Virtual I/O Server. All of this is done automatically for you during the migration.

- ► Use standard SAN configuration techniques to attach the same remote storage that you attached to the source Virtual I/O Server.
  - – The same remote LUNs must be available as hdisks on both the source and destination Virtual I/O Servers.
- ► Configure shared Ethernet adapters for each physical network that is configured on the mobile partition, just as you did on the source Virtual I/O Server.
- ► Create one VASI device for the Virtual I/O Server.
- ► Ensure the **Mover service partition** box is checked in the Virtual I/O Server partition properties.

Figure 5-11 shows the created and configured destination Virtual I/O Server. In the figure, the hdisk numbers on the destination Virtual I/O Server differ from those on the source Virtual I/O Server. The hdisk numbers may be different, but they are the same LUNs on the storage subsystem.
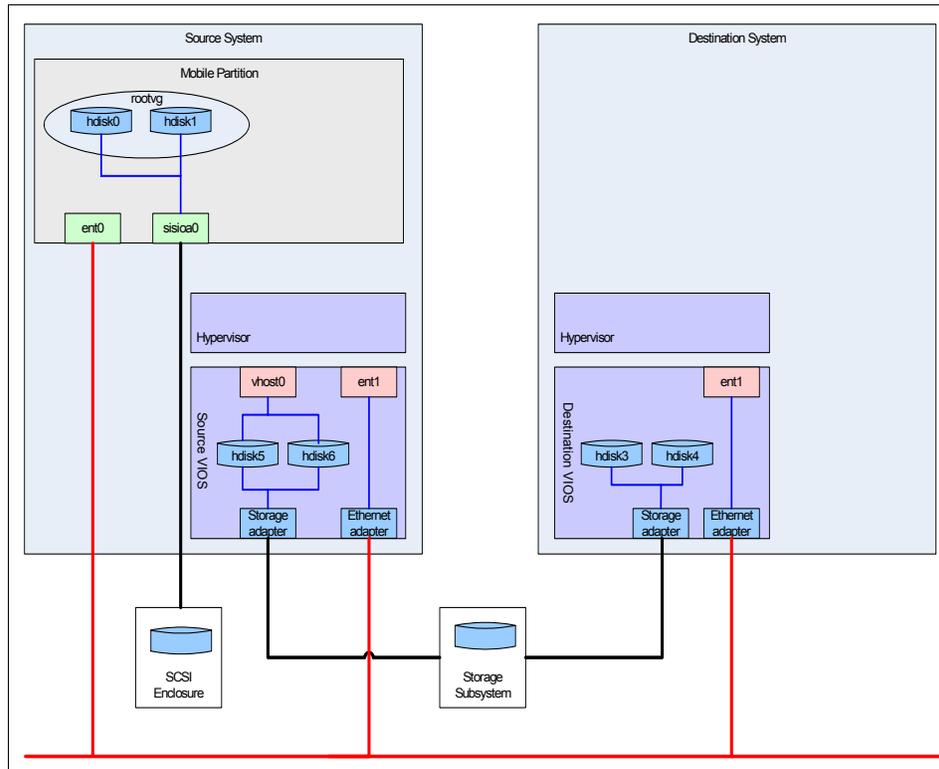
Figure 5-11   The destination Virtual I/O Server is created and configured

5.5.4 Configure storage on the mobile partition

Perform the following steps to switch over to using virtual storage devices on your mobile partition:

1. Add a virtual SCSI client adapter to the profile of the mobile partition. Ensure the virtual SCSI client adapter refers to the server adapter you created on the source Virtual I/O Server.

2. Use dynamic logical partitioning to add a virtual SCSI client adapter with the same properties from the previous step to the running mobile partition.

3. Configure the virtual SCSI devices on the mobile partition.

   a. Run the `cfgmgr` command on the mobile partition.

   b. Verify that the virtual SCSI adapters are in the Available state using the `lsdev` command.

      ```
      # lsdev -t IBM,v-scsi
      ```

   c. Verify that the virtual SCSI disks are in the Available state using the `lsdev` command.

      ```
      # lsdev -t vdisk
      ```

Figure 5-12 shows the configured storage devices on the mobile partition.
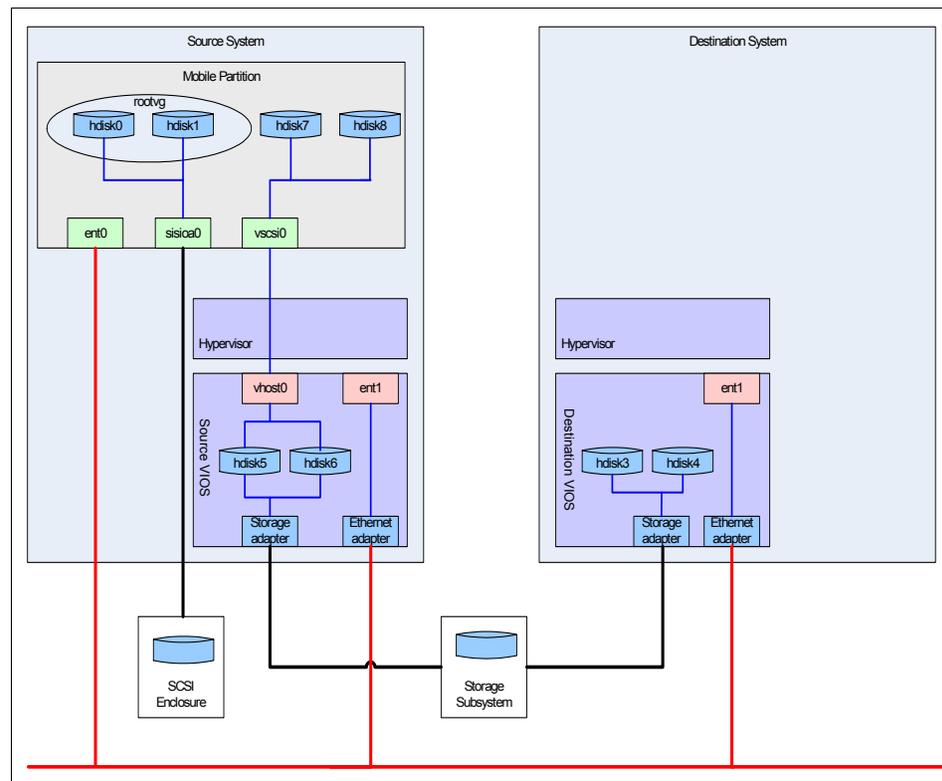
Figure 5-12   The storage devices are configured on the mobile partition

4. On the mobile partition, move rootvg from physical disks on to virtual disks.
   For example, assume hdisk0 and hdisk1 are the physical disks in rootvg, and
   that hdisk7 and hdisk8 are the virtual disks you created whose sizes are at
   least as large as hdisk0 and hdisk1:

   a. Extend rootvg on to virtual disks using the **extendvg** command.

      ```
      # extendvg rootvg hdisk7 hdisk8
      ```

      - Depending on the size of the disks, you may need to change the factor
        of the volume group with the **chvg** command in order to extend to the
        new disks. Do not do this unless the **extendvg** command fails.

        ```
        # chvg -t 10 rootvg
        ```

        Figure 5-13 shows rootvg extended on to the virtual disks.
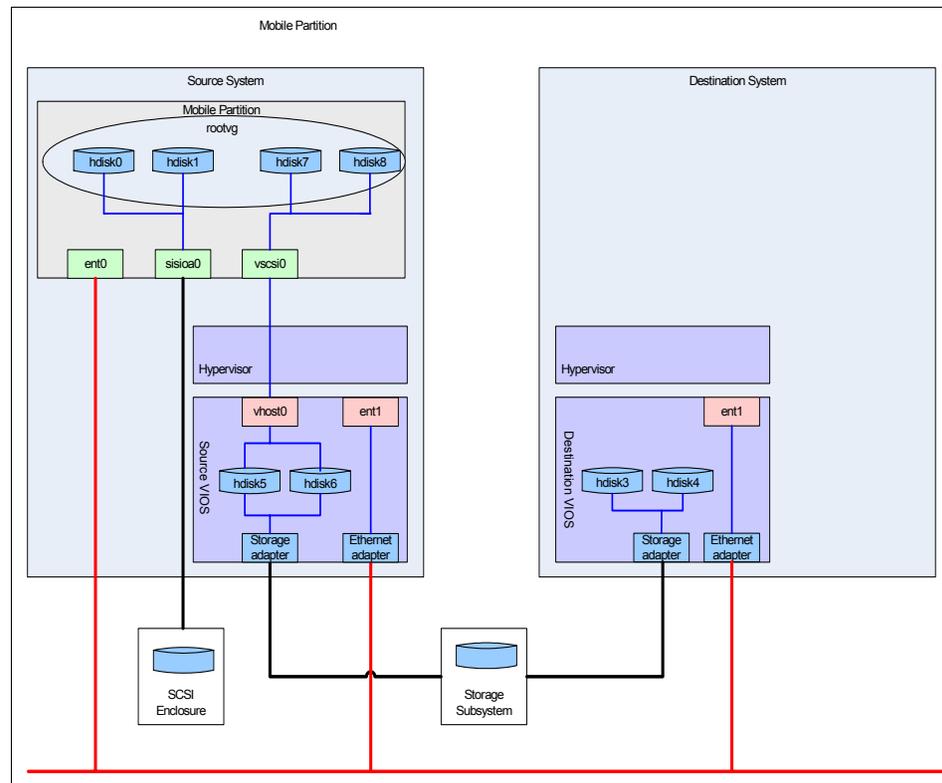
Figure 5-13   The root volume group extends on to virtual disks

b. Migrate physical partitions off the physical disks in rootvg on to the virtual disks in rootvg using the **migratepv** command.

```
# migratepv hdisk0 hdisk7
# migratepv hdisk1 hdisk8
```

c. Set the bootlist to a virtual disk in rootvg using the **bootlist** command.

```
# bootlist -m normal hdisk7 hdisk8
```

d. Run the **bosboot** command on a virtual disk in rootvg.

```
# bosboot -ad /dev/hdisk7
```

e. Remove physical disks from rootvg using the **reducevg** command.

```
# reducevg rootvg hdisk0 hdisk1
```

5. Repeat the previous step (excluding the **bootlist** command and the **bosboot** command) for all other volume groups on the mobile partition.

Figure 5-14 shows rootvg on the mobile partition now wholly on the virtual disks.
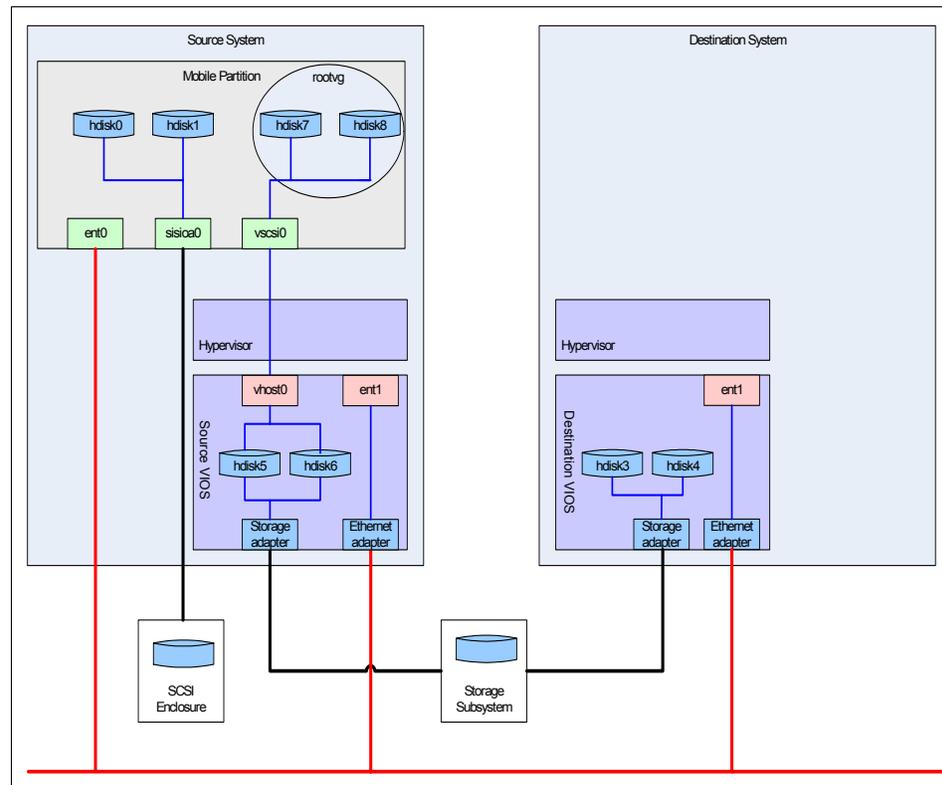
Figure 5-14   The root volume group of the mobile partition is on virtual disks only

5.5.5 Configure network on the mobile partition

Perform the following steps to configure the virtual network devices on the mobile partition:

1. Add virtual Ethernet adapters to the profile of your mobile partition.

   – One virtual adapter should be added for each physical network on the mobile partition.

   – Ensure the virtual Ethernet adapters use the shared Ethernet adapters you created on the source Virtual I/O Server.

2. Use dynamic logical partitioning to add virtual Ethernet adapters with the same properties from the previous step to the running mobile partition.

3. Configure the virtual Ethernet devices on the mobile partition.

   a. Run the **cfgmgr** command on the mobile partition to make the devices available.

   b. Verify that the virtual Ethernet adapters are in the Available state using the **lsdev** command:

   ```
   # lsdev -t IBM,l-la
   ```

Figure 5-15 shows the mobile partition with a virtual network device created.
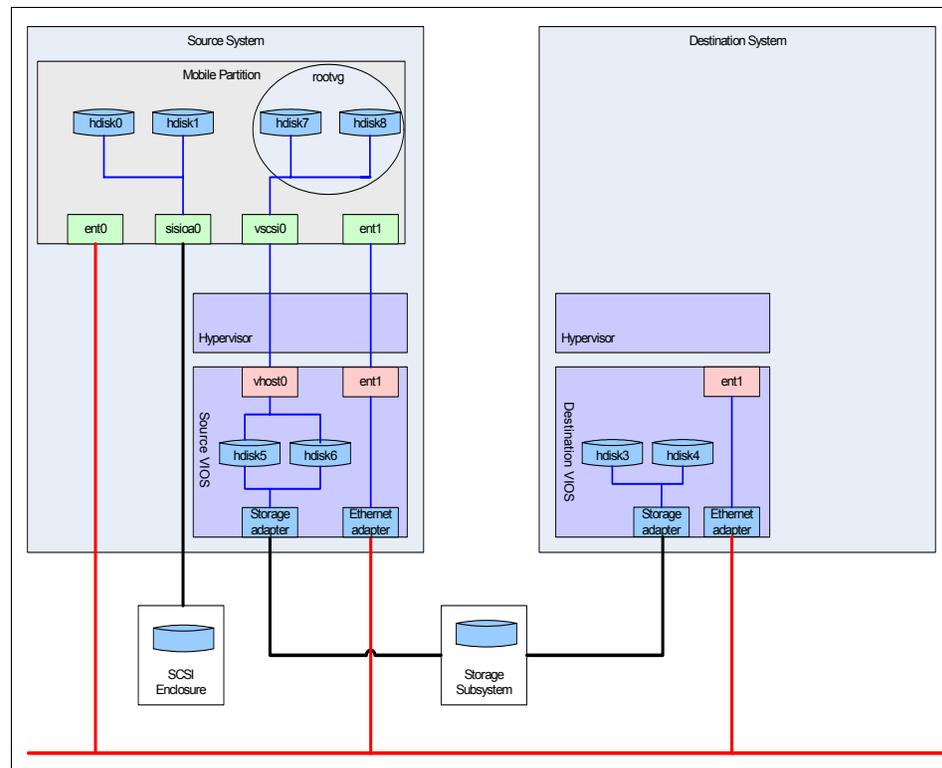
Figure 5-15   The mobile partition has a virtual network device created

Now that the virtual network adapters are configured, you must stop using the physical network adapters and begin using the virtual network adapters. To move to virtual networks on the mobile partition, there are a couple of options. Both procedures impact network connectivity differently. Understand how all running applications use the networks and take appropriate actions before proceeding.

Use new IP addresses
In order to follow this process, obtain a new IP address for each physical network that the mobile partition is using. Then perform the following steps:

1. Configure the virtual network interfaces on the mobile partition, using the new IP addresses that you obtained.

2. Verify network connectivity for each of the virtual network interfaces.

3. Unconfigure the physical network interfaces on the mobile partition.

Use existing IP addresses

In order to follow this process, record the network information for the physical
network interfaces on the mobile partition, and then perform the following steps:

1. Unconfigure the physical network interfaces on the mobile partition.

2. Configure the virtual network interfaces on the mobile partition, using the IP
   addresses previously used by the physical interfaces.

3. Verify network connectivity for each of the virtual interfaces.

Figure 5-16 shows the mobile partition using a virtual network, with its physical
network interface unconfigured.
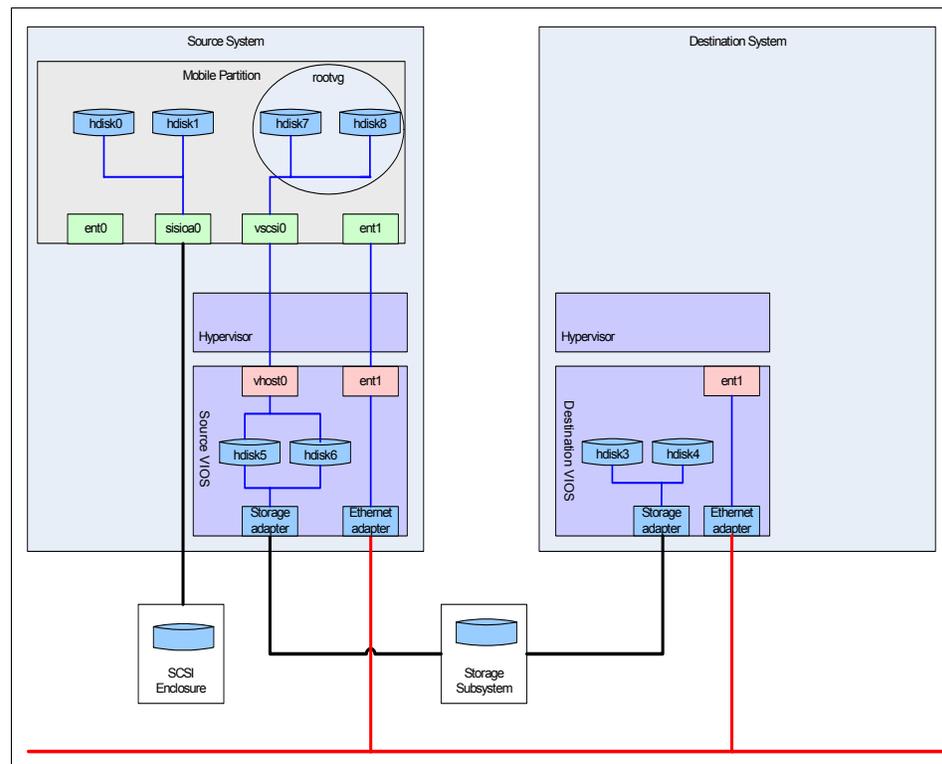

Figure 5-16   The mobile partition has unconfigured its physical network interface

5.5.6 Remove adapters from the mobile partition

There are two different procedures to remove adapters that cannot be migrated
from the mobile partition. If the mobile partition has any such adapters that are
marked as required, follow the first process. Otherwise, follow the second
process.

Process 1: Required adapters

Perform the following steps to remove the adapters from the mobile partition:

1. Remove all physical adapters (including HEA) from the profile of the mobile partition.

2. Remove all virtual serial adapters in slots 2 and above from the profile of the mobile partition.

3. Shut down the mobile partition.

4. Activate the mobile partition with the modified profile.

> **Note:** A reboot is not sufficient. The mobile partition must be shut down and activated with the modified profile.

Process 2: No required adapters

Perform the following steps to remove the adapters from the mobile partition:

1. At this point, no physical devices are in use on the mobile partition. Remove all physical devices, along with their children, using the **rmdev** command. For example, if the only physical devices in use are in slots pci0 and pci1, run the following commands to remove the physical devices:

```
# rmdev -R -dl pci0
# rmdev -R -dl pci1
```

2. Remove all physical adapters from the mobile partition using dynamic logical partitioning.

3. Remove all virtual serial adapters from slots 2 and above from the mobile partition using dynamic logical partitioning.

Figure 5-17 shows the mobile partition with only virtual adapters.
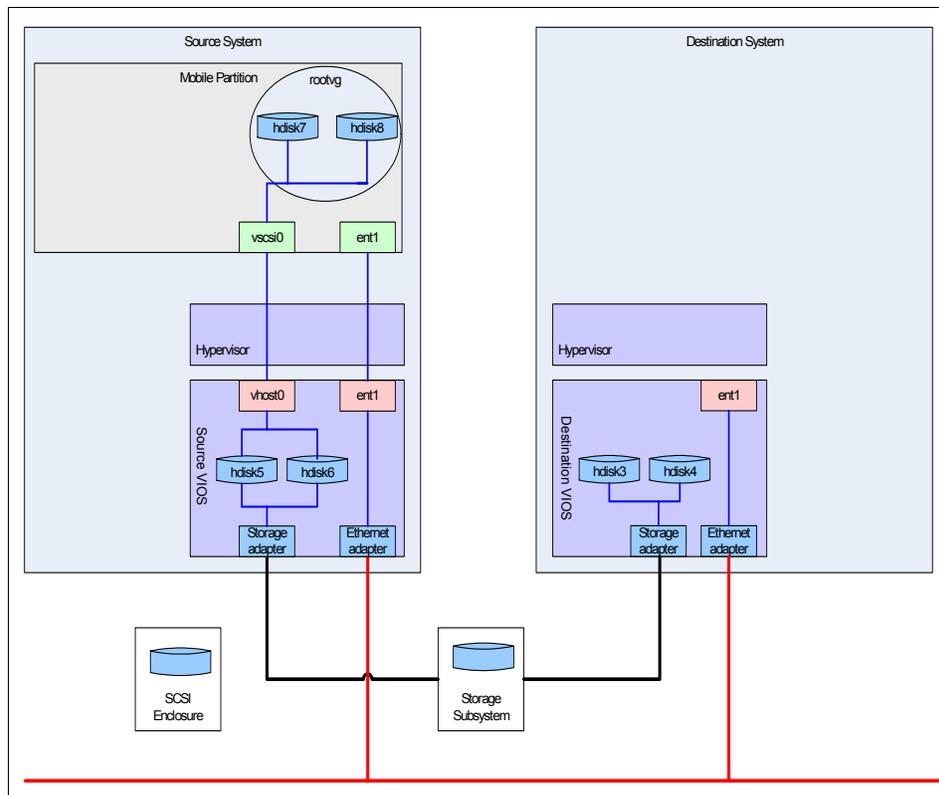
Figure 5-17   The mobile partition with only virtual adapters

5.5.7 Ready to migrate

The mobile partition is now ready to be migrated. Close any virtual terminals on the mobile partition, since they will lose connection when the partition migrates to the destination system. Virtual terminals can be reopened when the partition is on the destination system.

After the migration is complete, consider adding physical resources back to the mobile partition, if they are available on the destination system.

Note: The active mobile partition profile is created on the destination system without any references to any physical I/O slots that were present in your profile on the source system. Any other mobile partition profiles are copied unchanged.

Figure 5-18 shows the mobile partition migrated to the destination system.
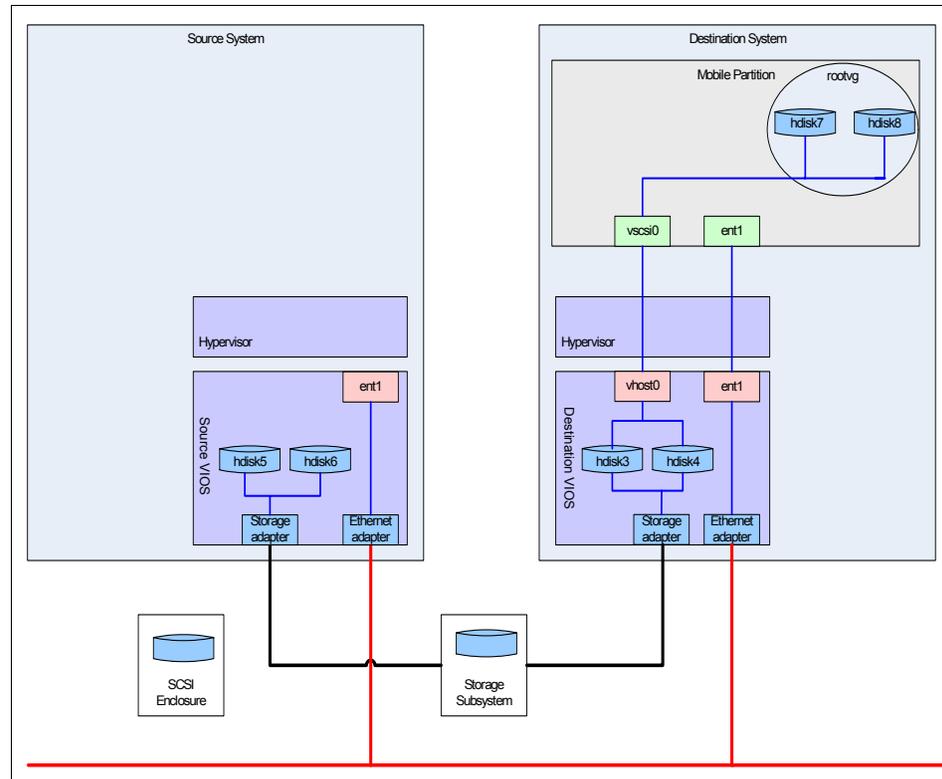
Figure 5-18   The mobile partition on the destination system

5.6 The command-line interface (CLI)

The HMC provides a command-line interface as well as an easy-to-use user interface for Live Partition Mobility. This CLI allows you to script frequently performed operations. This automation saves time and reduces the likelihood of errors.

There are two new HMC commands for Live Partition Mobility and five existing commands have been updated to support mobility. The new commands are `migrlpar` and `lslparmigr`. The modified commands are `lssyscfg`, `mksyscfg`, `chsyscfg`, `lshwres`, and `chhwres`.

Note: There is an existing command, `migrcfg`, that is used to push partition configuration data held on the HMC to a managed system. Despite its migr prefix, this operation is quite distinct from the Live Partition Mobility described in this publication.

The HMC commands can be launched either locally on the HMC or remotely using `ssh -l <hmc> <hmc_command>`.

Tip: Use the `ssh-keygen` command to create the public/private key-pair on your client. Then add these keys to the HMC user's key-chain using the `mkauthkeys --add` command on the HMC.

The new commands follow the HMC command conventions:

► Single character parameters are preceded by a single dash "-".

► Multiple character parameters are preceded by a double-dash "--".

► All filter and attribute names are lower case, with underscores joining words together, for example vios_lpar_id.

5.6.1 The migrlpar command

This new command is used to validate, initiate, stop, and recover a partition migration. The same command, syntax, and options are used for both active and inactive migrations. The HMC determines which type of migration to perform based on the state of the partition referenced in the command. The command syntax is:

```
migrlpar          -o m | r | s | v
                      -m <managed system>
                      [-t <managed system>]
                      -p <partition name> | --id <partitionID>
                      [-n <profile name>]
                      [-f <input data file> | -i <input data>]
                      [-w <wait time>]
                      [--force]
                      [-d <detail level>]
                      [-v]
                      [--help]
```

-o	The operation to perform m - validate and migrate r - recover s - stop v - validate
-m <managed system>	The source managed system's name.
-t <managed system>	The destination managed system's name.
-p <partition name>	The partition on which to perform the operation.
--id <partitionID>	The ID of the partition on which to perform the operation.
-n <profile name>	The name of the partition profile to be created on the destination.
-f <input data file>	The name of the file containing input data for this command. The format is: attr_name1=value,attr_name2=value,... or attr_name1=value1,value2,...
-i <input data>	The input data for this command, typically the virtual adapter mapping from source to destination or the destination shared-processor pool. This follows the same format as the input data file of the -f option.
-w <wait time>	The time, in minutes, to wait for any operating system command to complete.
--force	Force the recovery. This option should be used with caution.
-d <detail level>	The level of detail requested from operating system commands; values range from 0 (none) to 5 (highest).
-v	Verbose mode.
--help	Prints a help message.

Input data format

The data given in the file specified with the -f flag, or the data specified with -i, must be in comma-separated (CSV) format. These switches can be used with the migrate (-o m) and the validate (-o v) operations. The following attributes are supported: virtual_scsi_mappings, vlan_mappings, source_msp_name, source_msp_id, dest_msp_name, and dest_msp_id, shared_proc_pool_id, shared_proc_pool_name.

The data specified with the virtual_scsi_mappings attribute consists of one or more source virtual SCSI adapter to destination virtual SCSI adapter mappings in the format: client_virtual_slot_num/dest_vios_lpar_name/dest_vios_lpar_id.

Validate operation (-o v)

The `migrlpar -o v` command validates the proposed migration, returning a non-zero return code if the validate operation finds any configuration errors that will cause the migration to fail. Warnings not accompanied by an error do not cause the validate operation to fail. The command output is a list of errors and warnings of every potential or real problem that the HMC finds. The HMC does not stop the validation process at the first error; it continues processing as far as possible in an attempt to identify all problems that might invalidate the migration.

Migrate operation (-o m)

`migrlpar -o m` initiates a migration. The same command is used for inactive or active migrations; the HMC chooses the appropriate migration type based on the state of the given partition.

Stop operation (-o s)

If the migration is in a stoppable state, the `migrlpar -o s` command will halt the specified migration and roll back any changes. This command can only be executed on the HMC upon which the migration was started.

Recovery operation (-o r)

In the event of a lost connection or a migration failure, you can use the recovery operation to restore a partially migrated state using the `migrlpar -o r` command. Depending on what point in the migration the connection was lost, the recovery command either rolls-back the operation (undoes the changes on the destination system) or completes the migration.

Examples

To migrate the partition myLPAR from the system srcSystem to the destSystem using the default MSPs and adapter maps, use this command:

```
$ migrlpar -o m srcSystem -t destSystem -p myLPAR
```

In an environment with multiple mover service partitions on the source and destination, you can specify which mover service partitions to use in a validation or migration operation. The following command validates the migration in the previous example with specific mover service partitions. Note that you can use both partition names and partition IDs on the same command:

```
$ migrlpar -o v -m srcSystem -t destSystem -p myLPAR \
-i source_msp_id=2,dest_msp_name=S2_VIOS2
```

To specify the VLAN Virtual I/O Server mappings for a migration, use the -i or -f flags. In this example, we show the use of the partition id (--id):

```
$ migrlpar -o m -m srcSystem -t destSystem --id 4 \
-i vlan_mappings=1/S2_VIOS2/2
```

When the destination system has multiple shared-processor pools, you can stipulate to which shared-processor pool the moving partition will be assigned at the destination with either of the following commands:

```
$ migrlpar -o m  -m srcSystem -t destSystem -p myLPAR -i
"shared_proc_pool_id=1"
```

or

```
$migrlpar -o m -m srcSystem -t destSystem -p myLPAR -i
"shared_proc_pool_name="DefaultPool"
```

The capacity of the chosen shared-processor pool must be sufficient to accommodate the migrating partition otherwise the migration operation will fail.

The syntax to stop a partition migration is:

```
$ migrlpar -o s -m srcSystem -p MyLPAR
```

The syntax to recover a failed migration is:

```
$ migrlpar -o r -m srcSystem -p MyLPAR
```

You can use the --force flag on the **recover** command, but you should only do this when the partition migration fails, leaving the partition definition on both the source and destination systems.

5.6.2 The lslparmigr command

Use this new command to show the state of running migrations or to show the managed mover service partitions, Virtual I/O Servers, and adapter mappings that might be used for a partition migration. The syntax is shown below:

lslparmigr   -r lpar | msp | sys | virtualio
                          -m <managed system>
                          [-t <managed system>
                          --filter <filter data>
                          [-F [<input data file>] [--header]]
                          [--help]

-r   The type of resources for which to list information:
                          lpar - partition
                          msp - mover service partitions

sys- managed system (CEC)
virtualio - virtual I/O

-m <managed system> The source managed system's name.

-t <managed system> The destination managed system's name.

--filter <filter data>

Filters the data to be listed in CSV format.
The syntax is:
filter_name1=value,filter_name2=value,...
or

filter_name1=value1,value2,...
Valid filter names are: lpar_ids and lpar_names

The filters are mutually exclusive. This parameter is not valid with -r sys, optional with -r lpar and required with -r msp and -r virtualio. With -r msp and -r virtualio, exactly one partition name or ID must be specified and the partition must be an AIX or Linux partition.

-F [<attribute names>] Comma-separated list of the names of the attributes to be listed. If no attribute names are specified, then all attributes will be listed.

--header Prints a header of attribute names when -F is also specified.

--help Prints a help message.

Partition information (-r lpar)

The `lslparmigr -r lpar` command displays partition migration information. Without the -F flag, it lists the following attributes: lpar_name, lpar_id, migration_state, migration_type, source_sys_name, dest_sys_name, source_lpar_id, dest_lpar_id, source_msp_name, source_msp_id, dest_msp_name, and dest_msp_id.

Mover service partition information (-r msp)

The `lslparmigr -r msp` option displays the MSP-to-MSP relationship between the source and destination systems, that is, for each MSP on the source system, the command will display the mover service partition on the destination system it can communicate with. The following attributes are listed: source_msp_name, source_msp_id, dest_ip_names, and dest_msp_ids. If there are no MSPs on the source system, there will be no data.

System information (-r sys)

The `lslparmigr -r sys` command displays all partition migration information for a managed system. The following attributes are listed:
inactive_lpar_migration_capable, active_lpar_migration_capable,
num_inactive_migrations_supported, num_active_migrations_supported,
num_inactive_migrations_in_progress, and num_active_migrations_in_progress.

Virtual I/O Server information (-r virtualio)

The `lslparmigr -r virtualio` command displays information pertaining to the candidate destination Virtual I/O Servers. The command shows the possible and suggested mappings between the source virtual client adapters and the destination virtual server adapters for a given migration for both SCSI and VLAN virtual adapters.

Examples

The following examples illustrate how this command is used.

System migration information

To display the migration capabilities of a system, use the following syntax:

```
$ lslparmigr -r sys -m mySystem
```

which produces:

```
inactive_lpar_mobility_capable=1,num_inactive_migrations_supported=40,
num_inactive_migrations_in_progress=1,active_lpar_mobility_capable=1,
num_active_migrations_supported=40,num_active_migrations_in_progress=0
$
```

In this example, we can see that the system is capable of both active and inactive migration and that there is one inactive partition migration in progress. Using the -F flag produces the same information in a CSV format:

```
$ lslparmigr -r sys -m mySystem -F
```

which produces:

```
1,40,1,1,40,0
```

These are the same attribute values seen in the preceding example, without the attribute identifier. This format is appropriate for parsing or for importing in to a spreadsheet. Adding the --header switch prints column headers on the first line:

```
$ lslparmigr -r sys -m mySystem -F --header
```

which produces:

```
inactive_lpar_mobility_capable,num_inactive_migrations_supported,
num_inactive_migrations_in_progress,active_lpar_mobility_capable,
num_active_migrations_supported,num_active_migrations_in_progress
1,40,1,1,40,0
```

On a terminal, the header is printed on a single line.

If you are only interested in specific attributes, then you can specify these as options to the -F flag. For example, if you want to know just the number of active and inactive migrations in progress, you would use the command:

```
$ lslparmigr -r sys -m mySystem -F \
num_active_migrations_in_progress,num_inactive_migrations_in_progress
```

which produces:

```
0,1
```

Indicating that there are no active migrations and one active migration running. If you want a space instead of a comma to separate values, you must surround the attributes with double quotes.

Partition migration information

To show the migration information of the logical partitions of a managed system, use the **-r lpar** option:

```
$ lslparmigr -r lpar -m mySystem
```

which produces:

```
name=DEV,lpar_id=6,migration_state=Not Migrating
name=S2_VIOS2,lpar_id=2,migration_state=Not Migrating
name=S2_VIOS1,lpar_id=1,migration_state=Not Migrating
name=PROD,lpar_id=3,migration_state=Migration
Starting,migration_type=inactive,dest_sys_name=9117-MMA-SN10FFE0B-L9,
dest_lpar_id=65535
```

Here we can see that the system mySystem is hosting four partitions, DEV, S2_VIOS2, S2_VIOS1, and PROD. Of these, the PROD partition is in the Starting state of an inactive migration as indicated by the migration_state and migration_type attributes. At the point in time, the command was run, the migration the ID of the destination partition had not been chosen, as seen by the 65535 value for the dest_lpar_id parameter.

Use the -filter flag to limit the output to a given set of partitions with either the lpar_names or the lpar_ids attributes:

```
$ lslparmigr -r lpar -m mySystem --filter lpar_ids=3
```

which produces:

```
name=PROD,lpar_id=3,migration_state=Migration Starting,
migration_type=inactive,dest_sys_name=9117-MMA-SN10FFE0B-L9,
dest_lpar_id=7
```

Here the output information is limited to the partition with ID=3, which is the one performing the inactive migration. Here we can see that the dest_lpar_id has now been chosen.

You can use the -F switch to generate the same information in CSV format or to limit the output:

```
$ lslparmigr -r lpar -m mySystem --filter lpar_ids=3 -F
```

which produces:

```
PROD,3,Migration Starting,inactive,9117-MMA-SN10DD4AC-L10,
9117-MMA-SN10FFE0B-L9,3,7,,unavailable,,unavailable
```

Here the -F switch, without additional parameters, has printed all the attributes. In the above example, the last four fields of output pertain to the MSPs; because the partition in question is undergoing an inactive migration, no MSPs are involved and these fields are empty. You can use the --header flag with the -F flag to print a line of column headers at the start of the output.

Mover service partition information

The **-r msp** option shows the possible mover service partitions for a given migration:

```
$ lslparmigr -r msp -m srcSystem -t destSystem --filter lpar_names=TEST
```

which produces:

```
source_msp_name=S1_VIOS1,source_msp_id=1,
"dest_msp_names=S2_VIOS2,S2_VIOS1","dest_msp_ids=2,1"
source_msp_name=S1_VIOS2,source_msp_id=2,
"dest_msp_names=S2_VIOS2,S2_VIOS1","dest_msp_ids=2,1"
```

Here we can see that were we to move the partition TEST from srcSystem to destSystem, then:

▶ There are two mover service partitions on the source (S1_VIOS1 and S1_VIOS2).

▶ There are two mover service partitions on the destination (S2_VIOS1 and S2_VIOS2).

▶ If the migration uses S1_VIOS1 on the source, then either S2_VIOS1 or S2_VIOS2 could be used on the destination.

▶ If the migration uses S1_VIOS2 on the source, then either S2_VIOS1 or S2_VIOS2 could be used on the destination.

This gives four possible mover service partition combinations for the migration.

Virtual I/O Server

Use the -r virtualio option to display the possible virtual adapter mappings for a given migration. The -F flag can be used to format or limit the output, as shown below:

```
$ lslparmigr -r virtualio -m srcSystem -t destSystem \
--filter lpar_names=TEST -F suggested_virtual_scsi_mappings
```

which produces:

```
4/S2_VIOS2/2
```

This output indicates that were you to migrate the client partition called TEST from the srcSystem to destSystem, then the suggested virtual SCSI adapter mapping would be to map the client virtual adapter in slot 4 to the Virtual I/O Server called S2_VIOS2, which has an partition ID of 2, on the destination system.

5.6.3 Modified HMC commands

Five existing HMC commands have been modified to support Live Partition Mobility. These are `lssyscfg`, `mksyscfg`, `chsyscfg`, `lshwres`, and `chhwres`.

The lssyscfg command

`lssyscfg -r sys` displays two new attributes, active_lpar_migration_capable and inactive_lpar_migration_capable. These attributes can have a value of either 0 (incapable) or 1 (capable).

`lssyscfg -r lpar` displays the new msp and time_ref partition attributes on Virtual I/O Server partitions that are capable of participating in active partition migrations.

`lssyscfg -r prof` displays the new virtual_vasi_adapters profile attributes on Virtual I/O Server partitions that are capable of participating in active partition migrations.

The mksyscfg and chsyscfg commands

The `mksyscfg -r [lpar | prof]` and `chsyscfg -r prof` commands accept the new partition/profile attribute virtual_vasi_adapters. This attribute is valid for Virtual I/O Servers, and then only if the Virtual I/O Server partition is capable of participating in active partition migrations. The virtual_vasi_adapters attribute accepts any valid virtual slot number (only one slot number can be specified), or the value none.

The lshwres command

The `lshwres` command now displays the VASI adapters assigned to a Virtual I/O Server partition. The syntax is `lshwres -r virtualio --rsubtype vasi -m` `<managed system>` `[--level lpar]` `[--filter lpar_ids | lpar_names, slots]` `[-F [<attribute names]` `[--header]]`. This displays the lpar_name, lpar_id, and slot_num attributes.

The chhwres command

The **chhwres** command now supports the dynamic reconfiguration of VASI adapters into and out of Virtual I/O Server partitions capable of participating in active partition migrations. The syntax is **chhwres -r virtualio -m** <managed system> **-o** {**a** | **r**} {**-p** <partition name> | **--id** <partition ID>} **--rsubtype vasi** [**-s** <virtual slot number>] [**-w** <wait time>] [**-d** <detail level>] [**--force**].

5.6.4 A more complex example

Example 5-1 on page 150 provides a more complete example of how the Live Partition Mobility commands might be used. This script fragment moves all the migratable partitions from one system to another. The example assumes that two environment variables, SRC_SERVER and DEST_SERVER, point to the system to empty and the system to load, respectively.

The algorithm starts by listing all the partitions on SRC_SERVER. It then filters out any Virtual I/O Server partitions and partitions that are already migrating. For each remaining partition, it invokes a migration operation to DEST_SERVER. In this example, the migrations take place sequentially; it would be acceptable for them to be run in parallel provided there were no more than four concurrent active migrations per mover service partition. This is an exercise left to the reader.

How it works

The script starts by checking that both the source and destination systems are mobility capable. For this, it uses the new attributes given in the **lssyscfg** command. It then uses the **lslparmigr** command to list all the partitions on the system. It uses this list as an outer loop for the rest of the script. The program then performs a number of elementary checks:

► The source and destination must be capable of mobility.

  The **lssyscfg** command shows the mobility capability attribute.

► Only partitions of type aixlinux can be migrated.

  The script uses the **lssyscfg** command to ascertain the partition type.

► It would seem sensible to avoid migrating a partition that is already migrating.

  The script reuses the **lslparmigr** command for this.

► Validate the partition migration.

  For this, the script uses the **migrlpar -v** and checks the return code.

If all the checks pass, the migration is launched with the `migrlpar` command. The code snippet does some elementary error checking; if the `migrlpar` returns a non-zero value, a recovery is attempted using the `migrlpar -o r` command.

Example 5-1   Script fragment to migrate all partitions on a system

```
#
# Get the mobility capabilities of the source and destination systems
#
SRC_CAP=$(lssyscfg -r sys -m $SRC_SERVER \
          -F active_lpar_mobility_capable,inactive_lpar_mobility_capable)
DEST_CAP=$(lssyscfg -r sys -m $DEST_SERVER \
          -F active_lpar_mobility_capable,inactive_lpar_mobility_capable)


#
#
# Make sure that they are both capapble of active and inactive migration
#
   if [ $SRC_CAP = $DEST_CAP ] && [ $SRC_CAP = "1,1" ]
   then
#
# List all the partitions on the source system
#
      for LPAR in $(lslparmigr -r lpar -m $SRC_SERVER -F name)
      do
#
# Only migrate "aixlinux" partitions. VIO servers cannot be migrated
#
         LPAR_ENV=$(lssyscfg -r lpar -m $SRC_SERVER \
                     --filter lpar_names=$LPAR -F lpar_env)
         if [ $LPAR_ENV = "aixlinux" ]
         then
#
# Make sure that the partition is not already migrating
#
            LPAR_STATE=$(lslparmigr -r lpar -m $SRC_SERVER --filter lpar_names=$LPAR
-F migration_state)
            if [ "$LPAR_STATE" = "Not Migrating" ]
            then
#
# Perform a validation to see if there's a good chance of success
#
               migrlpar -o v -m $SRC_SERVER -t $DEST_SERVER -p $LPAR
               RC=$?
               if [ $RC -ne 0 ]
```

```
                then
                    echo "Validation failed. Cannont migrate partition $LPAR"
                else
#
# Everything looks good, let's do it...
#
                    echo "migrating $LPAR from $SRC_SERVER to $DEST_SERVER"
                    migrlpar -o m -m $SRC_SERVER -t $DEST_SERVER -p $LPAR
                    RC=$?
                    if [ $RC -ne 0 ]
                    then
#
# Something went wrong, let's try to recover
#
                        echo "There was an error RC = $RC . Attempting recovery"
                        migrlpar -o r -m $SRC_SERVER -p $LPAR
                        break
                    fi
                fi
            fi
        fi
    done
  fi
```

5.7 Migration awareness

A migration aware application is one that is designed to recognize and dynamically adapt to changes in the underlying system hardware after being moved from one system to another.

Most applications will not require any changes to work correctly and efficiently with Live Partition Mobility. Some applications may have dependencies on characteristics that change between the source and destination servers and other applications may adjust their behavior to facilitate the migration.

Applications that should probably be made migration aware include:

► Applications that use processor and memory affinity characteristics to tune their behavior because affinity characteristics may change as a result of migration. The externally visible behavior remains the same, but performance variations, for better or worse, may be observed due to different server characteristics.

Applications that use processor binding will maintain their binding to the same logical processors across migrations, but in reality the physical processors will have changed. Binding is usually done to maintain hot caches, but clearly the physical processor move will require a warming of the cache hierarchy on the destination system. This usually occurs very quickly and should not be visible to the users.

► Applications that are tuned for a given cache architecture, such as hierarchy, size, line-size, and associativity. These applications are usually limited to high-performance computing applications, but the just-in-time (JIT) compiler of the IBM Java virtual machine is also optimized for the cache-line size of the processor on which it was launched.

► Performance analysis, capacity planning, and accounting tools and their agents should also be made migration-aware because the processor performance counters may change between the source and destination servers, as may the processor type and frequency. Additionally, tools that calculate an aggregate system load based on the sum of the loads in all hosted partitions must be aware that a partition has left the system or that a new partition arrived.

► Workload managers

Actions that a migration-aware application might take include:

► Keep track of changes to system characteristics, such as cache-line size or serial numbers, and modify tuning or behavior accordingly.

► Terminate the application on the source system and restart it on the destination.

► Reroute workloads to another system.

► Clean-up system-specific buffers and logs.

► Refuse new incoming requests or delay pending operations.

► Increase time-out thresholds, such as the HACMP heartbeat.

► Block the sending of partition shutdown requests.

► Refuse a partition migration in the check phase to prevent a non-migratable application from being migrated.

5.8 Making applications migration-aware

Mobility awareness can be built in to an application using the standard AIX dynamic reconfiguration notification infrastructure. This infrastructure offers two different mechanisms for alerting applications to configuration changes.

▶ Using the SIGRECONFIG signal and the dynamic reconfiguration APIs

▶ Registering scripts with the AIX dynamic reconfiguration infrastructure

Using the SIGRECONFIG and dynamic reconfiguration APIs requires additional code in your applications. The DLPAR scripts allow you to add awareness to those applications for which you do not have the source code.

5.8.1 Migration phases

The DLPAR notification framework defines three operational phases: check, prepare, and post.

The check phase notification allows applications to signal their readiness to migrate. The check phase allows applications with root authority to refuse a migration.

The prepare phase notification alerts applications that the migration (or dynamic reconfiguration) is imminent. This allows applications to make any necessary steps to help with the process.

The post phase notification alerts applications that the migration (or dynamic reconfiguration) is complete. This allows applications to take any recovery steps to resume service on the destination system.

The check and prepare phases take place on the source system; the post phase occurs on the destination once the device tree and ODM have been updated to reflect the destination system configuration.

5.8.2 Making programs migration aware using APIs

Application programming interfaces are provided to make programs migration-aware. The SIGRECONFIG signal is sent to all applications at each of the three migration phases. Applications can watch (trap) this signal and use the DLPAR-API system calls to learn more about the operation in progress. Be aware that if your program does trap the SIGRECONFIG signal, it will be notified of all dynamic-reconfiguration operations and not just Live Partition Mobility events.

> **Note:** An application must not block the SIGRECONFIG signal and the signal must be handled in a timely manner. The dynamic LPAR and Live Partition Mobility infrastructure will wait a short period of time for a reply from applications. If after this time there has been no response, the system will assume all is well and proceed to the next phase. You can speed up a migration or dynamic reconfiguration operation by acknowledging the SIGRECONFIG event even if your application takes no action.

Applications must perform the following operations to be notified of a Live Partition Mobility operation:

1. Catch the SIGRECONFIG signal by using the sigaction() or sigwait() system calls. The default action is to ignore the signal.

2. Control the signal mask of at least one of the application's threads and the priority of the handling thread such that the signal can be delivered and handled promptly.

3. The signal handler uses the dr_reconfig() system call to determine the nature of the reconfiguration event as well as other pertinent information. For the check phase, the application should pass DR_RECONFIG_DONE to accept a migration or DR_EVENT_FAIL to refuse. Only applications with root authority may refuse a migration.

The dr_reconfig() system call has been modified to support partition migration. The returned dr_info structure now includes the following bit-fields:

► migrate

► partition

These are for the new migration action and the partition object that is the object of the action.

The code snippet in Example 5-2 gives an outline of how this might be used. This
code would run in a signal-handling thread.

Example 5-2   SIGRECONFIG signal-handling thread

```
#include <signal.h>
#include <sys/dr.h>
   :
   :
   struct dr_info  drInfo;       // For event-related information
   struct sigset_t signalSet;    // The signal set to wait on
   int             signalId;     // Identifies  signal was received
   int             reconfigFlag  // For accepting or refusing the DR
   int             rc;           // return code

   // Initialise the signal set
   SIGINITSET(signalSet);

   // Add the SIGRECONFIG to the signal set
   SIGADDSET(signalSet, SIGRECONFIG);

   // loop forever
   while (1) {
      // Wait on signals in signal set
      sigwait(&signalSet, &signalId);
      if (signalID == SIGRECONFIG) {
         if (rc = dr_reconfig(DR_QUERY, &drInfo)) {
            // handle the error
         } else {
            if (drInfo.migrate) {
               if {drInfo.check) {
                  /*
                   * If migration OK  reconfigFlag = DR_RECONFIG_DONE
                   * If migration NOK reconfigFlag = DR_EVENT_FAIL
                   */
                  rc = dr_reconfig(reconfigFlag, &drInfo);
               } else if (drInfo.pre) {
                  /*
                   * Prepare the application for migration
                   */
                  rc = dr_reconfig(DR_RECONFIG_DONE, &drInfo);
               } else if (drInfo.post) {
                  /*
                   * We're being woken up on the destination
                   * Check new environment and resume normal service
```

```
                    */
                } else {
                    // Handle the error cases
                }
            } else {
                // It's not a migration.  Handle or ignore the DR
        }
    }
}
```

You can use the sysconf() system call to check the system configuration on the
destination system. The _system_configuration structure has been modified to
include the following fields:

icache_size   Size of the L1 instruction cache

icache_asc   Associativity of the L1 instruction cache

dcache_size   Size of the L1 data cache

dcache_asc   Associativity of the L1 data cache

L2_cache_size   Size of the L2 cache

L2_cache_asc   Associativity of the L2 cache

itlb_size   Instruction translation look-aside buffer size

itlb_asc   Instruction translation look-aside buffer associativity

dtlb_size   Data translation look-aside buffer size

dtlb_asc   Data translation look-aside buffer associativity

tlb_attrib   Translation look-aside buffer attributes

slb_size   Segment look-aside buffer size

These fields are updated once the partition has arrived at the destination system
to reflect the underlying physical processor characteristics. In this fashion,
applications that are moved from one processor architecture to another can
dynamically adapt themselves to their execution environment. All new processor
features, such as the SIMD and decimal floating point instructions, are exposed
through the _system_configuration structure and the lpar_get_info() system call.

The lpar_get_info() returns two new capabilities, defined in <sys/dr.h>:

LPAR_INFO1_MSP_CAPABLE   If the partition is a Virtual I/O Server
                                    partition, the mover service partition
                                    capability is indicated.

LPAR_INFO1_PMIG_CAPABLE   Indicates if the partition is capable of
                                    migration.

5.8.3 Making applications migration-aware using scripts

Dynamic reconfiguration scripts allow you to cleanly quiesce and restart your applications over a migration. You can register your own scripts with the dynamic reconfiguration infrastructure using the `drmgr` command. This copies them to a private repository, the default location of which is /usr/lib/dr/scripts/all.

The scripts can be implemented in any interpreted (scripted) or compiled language. The `drmgr` command is detailed in the IBM InfoCenter at:

http://publib.boulder.ibm.com/infocenter/pseries/v5r3/topic/com.ibm.aix .cmds/doc/aixcmds2/drmgr.htm.

The syntax of the dynamic reconfiguration scripts is:

`[env_variable1=value ...] scriptname command [param1 ...]`

The input variables are set as environment variables on the command line, followed by the name of the script to be invoked and any additional parameters. Live Partition Mobility introduces four new commands, as shown in Table 5-1.

Table 5-1   Dynamic reconfiguration script commands for migration

Command and parameter	Description
checkmigrate <resource>	Used to indicate whether the a migration should continue or not. A script might indicate that a migration should not continue if the application is dependent upon an invariable execution environment. The script is called with this command at the check-migration phase.
premigrate <resource>	At this point the migration will be initiated. The script can reconfigure or suspend an application to facilitate the migration process. The script is called with this command at the prepare-migration phase.
postmigrate <resource>	This command is called after migration has completed. The script can reconfigure or resume applications that were changed or suspended in the prepare phase. The script is called with this command in the post-migration phase.
undopremigrate <resource>	If an error is encountered during the check phase, the script is called with this command to roll back any actions that might have been taken in the `checkmigrate` command in preparation for the migration.

In addition to the new script commands, there is a new pmig resource type to indicate a partition migration operation. The `register` command of your DLPAR

scripts can choose to handle this new resource type. A script supporting partition migration should write out the name-value pair DR_RESOURCE=pmig when it is invoked with the **register** command. A DLPAR script may be registered that only supports partition migration. There are no new environment variables passed to the DLPAR scripts for Live Partition Mobility support.

The code outline shown in Example 5-3 shows a Korn shell script that detects the partition migration reconfiguration events. For this example, the script simply logs the called command to a file.

Example 5-3   Outline Korn shell DLPAR script for Live Partition Mobility

```
#!/usr/bin/ksh

if [[ $# -eq 0 ]]
then
    echo "DR_ERROR=Script usage error"
    exit 1
fi

ret_code=0
command=$1

case $command in
    scriptinfo )
        echo "DR_VERSION=1.0"
        echo "DR_DATE=27032007"
        echo "DR_SCRIPTINFO=Parition migration test script"
        echo "DR_VENDOR=IBM"
        echo "SCRIPTINFO" >> /tmp/migration.log;;

    usage )
        echo "DR_USAGE=$0 command [parameter]"
        echo "USAGE" >> /tmp/migration.log;;

    register )
        echo "DR_RESOURCE=pmig";;
        echo "REGISTER" >> /tmp/migration.log;;

    checkmigrate )
        echo "CHECK_MIGRATE" >> /tmp/migration.log;;

    premigrate )
        echo "PRE_MIGRATE" >> /tmp/migration.log

    postmigrate )
```

```
        echo "POST_MIGRATE" >> /tmp/migration.log;;

    undocheckmigrate )
        echo "UNDO_CHECK_MIGRATE" >> /tmp/migration.log;;

    * )
        echo "*** UNSUPPORTED *** : $command" >> /tmp/migration.log;;
        ret_code=10;;
esac

exit $ret_code
```

If the file name of the script is migrate.sh, then you would register it with the dynamic reconfiguration infrastructure using the following command.

```
# drmgr -i ./migrate.sh
```

Use the **drmgr -l** command to confirm script registration, as shown in Example 5-4.

Example 5-4   Listing the registered DLPAR scripts

```
# drmgr -l
DR Install Root Directory: /usr/lib/dr/scripts
Syslog ID: DRMGR
-----------------------------------------------------------
/usr/lib/dr/scripts/all/migrate.sh               Parition migration test script
        Vendor:IBM,     Version:1.0,    Date:27032007
        Script Timeout:10,      Admin Override Timeout:0
        Memory DR Percentage:100
        Resources Supported:
                Resource Name: pmig      Resource Usage: /usr/lib/dr/scripts/all/migrate.sh command
[parameter]
-----------------------------------------------------------
```

In this example, you can see the output from the **scriptinfo**, **register**, and **usage** commands of the shell script.

5.9 Making kernel extension migration aware

Kernel extensions can register to be notified of migration events. The notification mechanism uses the standard dynamic reconfiguration mechanism - the reconfig_register() kernel service. The service interface signature is:

```
int reconfig_register_ext(handler, actions, h_arg, h_token, name)
int (*handler)(void*, void*, long long action, void* dri);
long long actions;
void* h_arg;
ulong *h_token;
char* name;
```

The actions parameter supports the following values for mobility awareness:

► DR_MIGRATE_CHECK

► DR_MIGRATE_PRE

► DR_MIGRATE_POST

► DR_MIGRATE_POST_ERROR

The interface to the handler is:

```
int handler(void* event, void* h_arg, long long action, void*
resource_info);
```

The action parameter indicates the specific reconfiguration operation being performed, for example, DR_MIGRATE_PRE. The resource_info parameter maps to the following structure for partition migration:

```
struct dri_pmig {
        int version;
        int destination_lpid;
        long long streamid
}
```

At the time of writing, the version number is 1. The version number will be changed if additional parameters are added to this structure. The destination_lpid and streamid fields are not available for the check phase.

The interfaces to the reconfig_unregister() and reconfig_complete() kernel services are not changed by Live Partition Mobility.

6

Migration status

This section contains topics related to migration status and recovery procedures to be followed when errors are found during the migration of a logical partition. A good knowledge of Live Partition Mobility prerequisites and actions is assumed.

The following subjects are covered:

► 6.1, "Progress and reference code location" on page 162

► 6.2, "Recovery" on page 164

6.1 Progress and reference code location

Live Partition Mobility is driven by the HMC. The HMC has knowledge of the status of all partition migrations and provides the latest reference code for each logical partition.

In order to see the migration status on the GUI, open **Systems Management** in the navigation area, click **Servers**, and select the managed system in which you are interested. A system status and reference code is provided for each logical partition. For example, in Figure 6-1 the QA partition is undergoing an active migration while an inactive migration is affecting the TRAIN partition, and for both partitions the latest reference code is displayed.

Figure 6-1   Partition reference codes

The same information can be obtained from the HMC command-line interface using the `lsrefcode` and `lslparmigr` commands. See 5.6, "The command-line interface (CLI)" on page 138 for details.

Reference codes describe the progress of the migration. You can find a description of reference codes in "SRCs" on page 184. When the reference code represents an error, a migration recovery procedure may be required.

After a migration is issued on the HMC GUI, a progress window is provided similar to the one shown in Figure 6-2. The percentage indicates the completion of memory state transfer during an active migration. In the case of an inactive migration, there is no memory management and the value is zero.
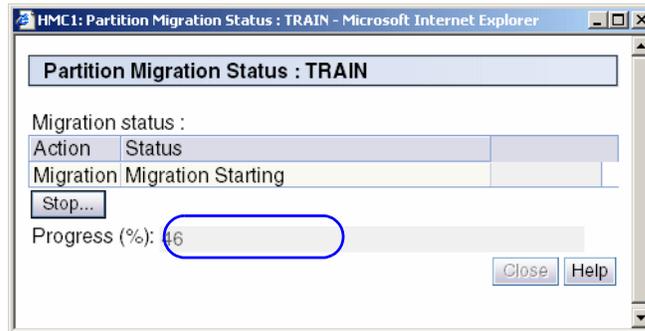
Figure 6-2   Migration progress window

During an inactive migration, only the HMC is involved, and it holds all migration information.

An active migration requires the coordination of the mobile partition and the two Virtual I/O Servers that have been selected as mover service partitions. All these objects record migration events in their error logs. You can find a description of partition related error logs in "Operating system error logs" on page 186.

The mobile partition records the start and the end of the migration process and you can extract the data with the **errpt** command, as shown in Example 6-1.

Example 6-1   Migration log on mobile partition

```
[Train:/]# errpt
IDENTIFIER TIMESTAMP  T C RESOURCE_NAME  DESCRIPTION
A5E6DB96   0329143307 I S pmig           Client Partition Migration Completed
08917DC6   0329143307 I S pmig           Client Partition Migration Started
...
```

Migration information is recorded also on the Virtual I/O Servers that acted as a mover service partition. It can be retrieved using the **errlog** command.

Example 6-2 show the data available on the source mover service partition. The first event in the log states when the mobile partition execution has been suspended on the source system and has been activated on the destination system, while the second records the successful end of the migration.

Example 6-2   Migration log on source mover service partition

```
$ errlog
IDENTIFIER TIMESTAMP  T C RESOURCE_NAME  DESCRIPTION
3EB09F5A   0329143407 I S Migration      Migration completed successfully
6CB10B8D   0329143407 I S                Client partition suspend issued
...
```

On the destination mover service partition, the error log registers only the end of the migration, as shown in Example 6-3.

Example 6-3   Migration log on destination mover service partition

```
$ errlog
IDENTIFIER TIMESTAMP  T C RESOURCE_NAME  DESCRIPTION
3EB09F5A   0329143407 I S Migration      Migration completed successfully
...
```

The error logs on the mobile partition and the Virtual I/O Servers also record events that prevent the migration from succeeding, such as user interruption or network problems. They can be used to trace all migration events on the system.

6.2 Recovery

Live Partition Mobility is designed to verify if a requested migration can be executed and to monitor all migration process. If a running migration cannot be completed, a roll back procedure is executed to undo all configuration changes applied.

A partition migration may be prevented from running for two main reasons:

- ► The migration is not valid and does not meet prerequisites.
- ► An external event prevents a migration component from completing its job.

The migration validation described in 4.4.1, "Performing the validation steps and eliminating errors" on page 97 takes care of checking all prerequisites. It can be explicitly executed at any moment and it does not affect by any means the mobile partition.

It is recommended to perform a validation before requesting any migration. The migration process, however, performs another validation right before starting any configuration changes.

Once the inactive or active migration begins, the HMC manages the configuration changes and monitors the status of all involved components. If any error occurs, it automatically starts recovery actions.

When the HMC cannot perform a recovery, administrator intervention is required to perform problem determination and issue final recovery steps. This situation may occur when the HMC cannot contact a migration component (for example, the mobile partition, a Virtual I/O Server, or a system service processor) due to a network problems or an operator error. After a timeout, an error message is provided requesting a recovery.

When a recovery is required, the mobile partition name may appear on both the source and the destination system. The partition is either powered down (inactive migration) or really working only on one of the two systems (active migration). Configuration cleanup is made during recovery.

While a mobile partition requires a recovery, its configuration cannot be changed to prevent any attempt to modify its state before its state is returned to normal operation. It is not possible to activate the same partition on two systems.

Recovery is performed by selecting the migrating partition and then selecting **Operations** → **Mobility** → **Recover**, as shown in Figure 6-3.
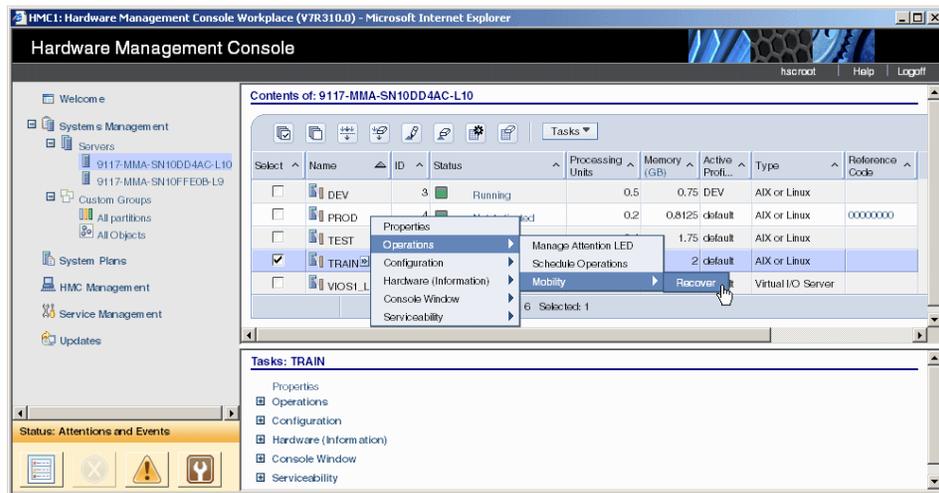
Figure 6-3   Recovery menu

A pop-up box appears similar to the one shown in Figure 6-4, requesting recovery confirmation. Click the **Recover** button to start a recovery. There is also a **Force recover** check box that should be used only when the HMC cannot contact one of the migration components that require a new configuration or if the migration has been stated by another HMC. The check box should be selected only after a normal recovery does not succeed.
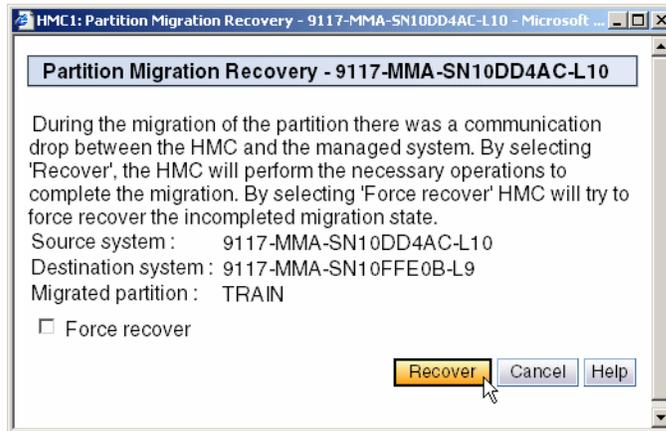
Figure 6-4   Recovery pop-up box

The same actions performed on the GUI can be executed using the `migrlpar` command on the HMC's command line. See 5.6, "The command-line interface (CLI)" on page 138 for details.

After a successful recovery, the partition returns to normal operation state and changes to its configuration are allowed. If the migration is executed again, the validation phase will detect the component that prevented the migration and select alternate elements or a validation error will be provided.

6.3 A recovery example

As an example, we have deliberately created a network outage during an active partition migration.

During an active migration, there is a partition state transfer through the network between the source and destination mover service partitions. The mobile partition keeps running on the source system while its state is copied on the destination system, then it is briefly suspended on the source and immediately reactivated on the destination.

We unplugged the network connection of one mover service partition in the middle of a state transfer. We had to perform several tests in order to create this scenario because the migration on the partition (2 GB of memory) was extremely fast.

On the HMC GUI, the migration process fails with an error message similar to Figure 6-5, suggesting you retry the operation later.
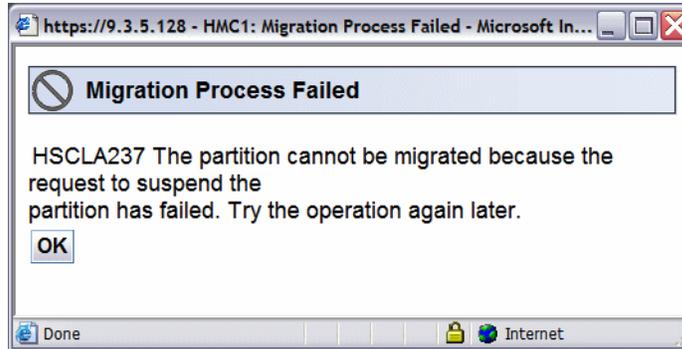
Figure 6-5   Error message after an active migration interrupted by network outage

Since the migration stopped in the middle of the state transfer, the partition configuration on the two involved systems is kept in the migrating status, waiting for the administrator to perform problem identification and decide how to continue.

On the HMC, the status of the migrating partition, TRAIN, is present in both systems, while it is active only on the source system. On the destination system, only the shell of the partition is present. The situation can be easily seen by opening the **Systems Management** menu in the navigation area, clicking **Custom Groups** and then **All partitions**. In the content area, a situation similar to Figure 6-6 is shown.

Figure 6-6   Interrupted active migration status

The applications running on the partition have not been affected by the network outage and are running on the source system. The only visible effect is on the partition's error log that shows the start and the abort of the migration, as described in Example 6-4. No action is required on the partition.

Example 6-4   Migrating partition's error log after aborted migration

```
[Train]# errpt
IDENTIFIER TIMESTAMP  T C RESOURCE_NAME  DESCRIPTION
5E075ADF   0403150307 I S pmig           Client Partition Migration Aborted
08917DC6   0403150207 I S pmig           Client Partition Migration Started
```

Both Virtual I/O Servers, used as a mover service partition, have recorded the event in their error log. On the Virtual I/O Server, where the cable was unplugged, we see both the physical network error and the mover service partition communication error, as described in Example 6-5.

Example 6-5   Mover service partition with network outage

```
$ errlog
IDENTIFIER TIMESTAMP  T C RESOURCE_NAME  DESCRIPTION
427E17BD   0403151907 P S Migration      Migration aborted: MSP-MSP connection do
0B41DD00   0403151707 I H ent4           ADAPTER FAILURE
...
```

The other Virtual I/O Server only shows the communication error of the mover service partition, because no physical error has been created, as described in Example 6-6.

Example 6-6   Mover service partition with communication error

```
$ errlog
IDENTIFIER TIMESTAMP  T C RESOURCE_NAME  DESCRIPTION
427E17BD   0403152107 P S Migration      Migration aborted: MSP-MSP connection do
...
```

In order to recover from an interrupted migration, you must select the mobile partition and select **Operations** → **Mobility** → **Recover**, as shown in Figure 6-3 on page 165.

A pop-up box similar to the one shown in Figure 6-4 on page 166 appears. Click the **Recover** button and the partition state is cleaned up; the TRAIN partition is present only on the source system where it is running, and it is removed on the destination system, where it has never been executed.

Once the network outage is resolved, the migration can be issued again. Wait for the RMC protocol to reset communication between the HMC and the Virtual I/O Server that had the network cable unplugged.

Integrated Virtual Ethernet for Live Partition Mobility

The Integrated Virtual Ethernet adapter is a new feature of the POWER6-based System p server. It provides advanced shared Ethernet ports between partitions in the same system.

This section describes how to configure and use the Integrated Virtual Ethernet adapter in the Live Partition Mobility environment.

A dedicated paper on the adapter is available at:

http://www.redbooks.ibm.com/abstracts/redp4340.html?Open

Integrated Virtual Ethernet adapter overview

A Integrated Virtual Ethernet adapter (IVE) is an integrated physical Ethernet adapter feature that is attached directly to the GX+ bus on a managed system. An IVE provides physical dual or quad 1 Gbps Ethernet (copper) ports, or dual 10 Gbps (fibre) ports.

A component of the IVE feature is named the HEA.

An HEA also provides an advanced shared Ethernet function, and each partition on the same system can connect directly to an IVE and access external networks through the IVE without configuring any Virtual I/O Servers.

To connect a partition to an HEA, you must create a logical host Ethernet adapter (LHEA) for the partition. An LHEA is a proxy of an HEA on a partition. An LHEA defines the resources that the partition can use on the HEA. Each logical partition can have at most one LHEA for each HEA on the system. Each LHEA can have one or more logical ports, and each logical port can connect to a physical port on the HEA. Each logical port has a unique MAC address.

Dual port 10 Gbps and quad port 1 Gbps HEA provide two port groups, and dual port 1 Gbps HEA provides one. Each port group can have 16 logical ports, up to 32 logical ports can be defined per HEA.

You can optionally connect logical ports on an LHEA with virtual LANs on your system. This allows the partition to communicate with other partitions on the same system and with external networks through a single logical adapter. You can connect a logical port with all of the virtual LANs on the system, or you can specify one or more virtual LANs to which the logical port can connect.

Figure A-1 shows a logical representation of an HEA adapter.
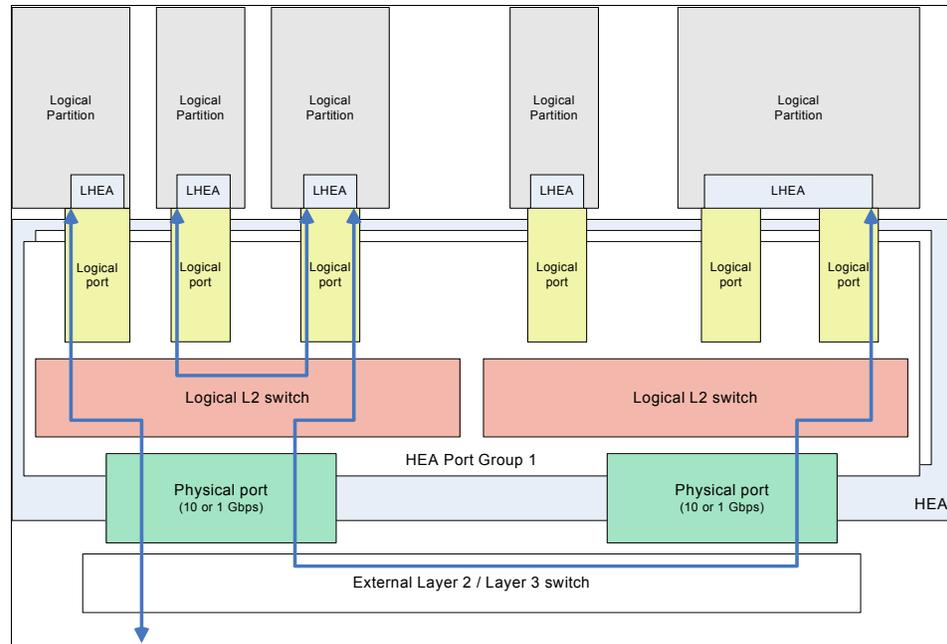
Figure A-1   Integrated Virtual Ethernet adapter overview

Requirements

The following sections discuss the requirements to use the HEA adapter with Live Partition Mobility.

Mobile partition

A partition with an LHEA cannot be migrated. If the mobile partition has one or more LHEAs, it cannot participate in an active or an inactive partition migration. Before migration, you must delete the LHEA definition from the mobile partition and the partition profile.

Shared Ethernet adapter on a Virtual I/O Server

If you use an LHEA as the physical adapter of an SEA, you must set the HEA into promiscuous mode. You can set the promiscuous mode by using the HMC. See "Setting properties of an IVE adapter" on page 174 about how to set the promiscuous mode.

If an HEA is in promiscuous mode, no other partitions may use that physical port.

Setting properties of an IVE adapter

You can set the properties of each IVE physical port. Any change that you make in the IVE port properties affects all partitions that use the IVE.

1. In the navigation pane on the HMC, expand **Systems Management** and click **Servers**.

2. In the Contents pane (the top right of the Hardware Management Console Workplace), select the system for which you want to set the properties of the IVE.

3. Click **View pop-up menu** and select **Hardware (Information)** →
   Adapters → **Host Ethernet**, as shown in Figure A-2.
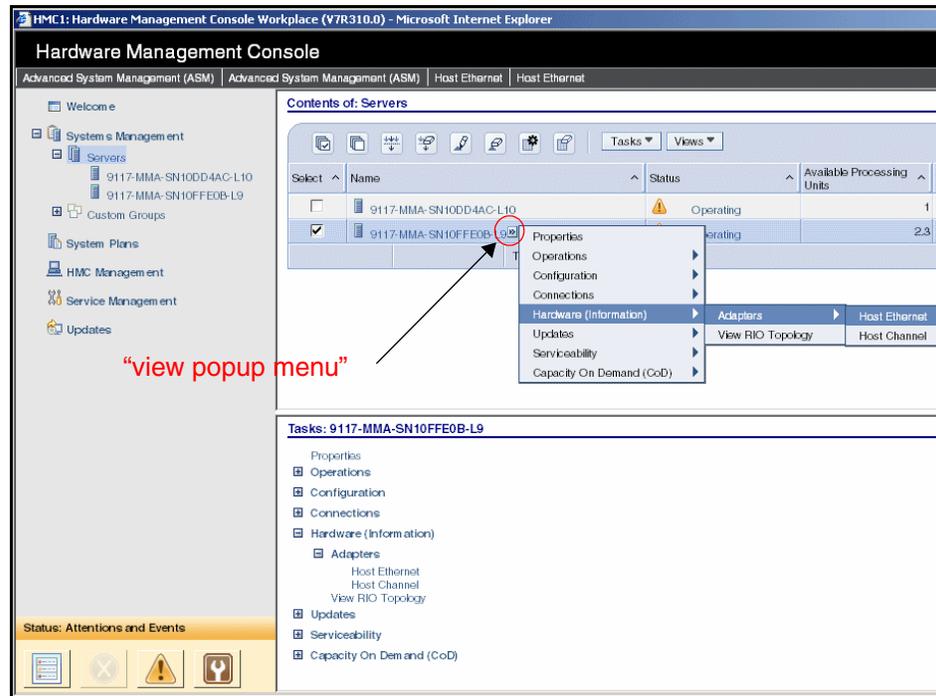
Figure A-2   HMC menu to set properties of a Integrated Virtual Ethernet adapter

4. Choose a Physical Location Code, select a physical port of the IVE for which you want to set properties, and then press the **Configure...** button in the Integrated Virtual Ethernet adapters window (Figure A-3).

The system shown in Figure A-3 has an IVE with dual 1 Gbps ports.
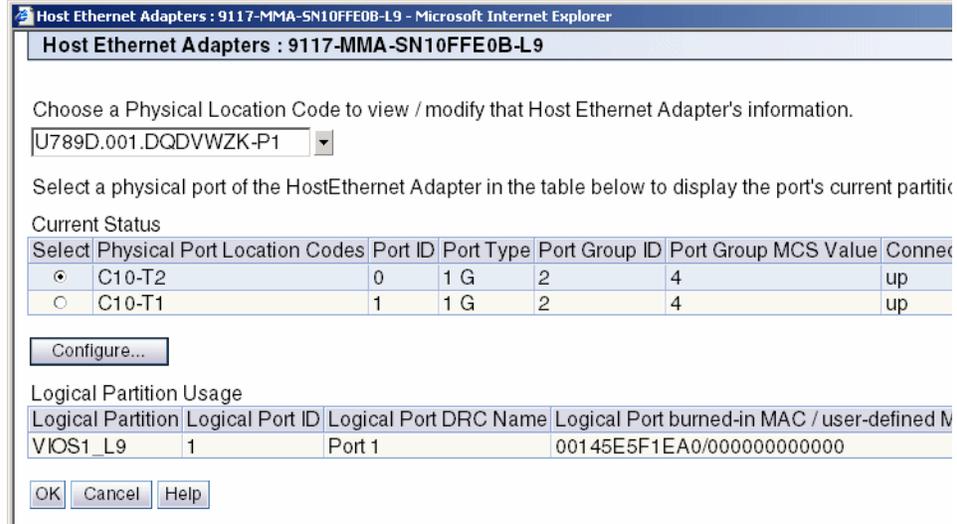
Figure A-3   Host Ethernet Adapters window

5. Set the properties of the IVE port in the HEA Physical Port Configuration window, and press **OK** (Figure A-4).
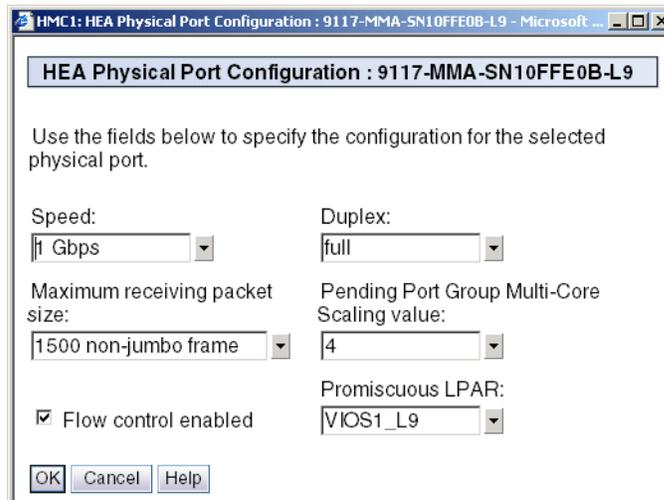
Figure A-4   HEA Physical Port Configuration window

– Pending Port Group Multi-Core Scaling value

Receive side distribution of packets and interrupts to multiple processor cores is accomplished through a mechanism called Multi-Core Scaling (MCS). When this feature is used, multiple default queues per logical port can be configured. The queue is determined according to a hash of the 6-tuple contained in the header of the incoming packet.

Increasing the MCS value (above 1) will effectively reduce the number of logical ports that are available on that port group, as shown in Table A-1. The MCS value chosen depends on how many logical ports are needed and how many processors are present.

Table A-1   Maximum logical ports according to MCS value

MCS value	1	2	4	8	16
Maximum logical ports allowed per port group	16	8	4	2	1

– Promiscuous LPAR

You can set this property on one partition on each physical port. If you set this property on any partition, that partition receives all the multicast and broadcast traffic.

This mode must be granted to the Virtual I/O Server partition if it has a shared Ethernet Adapter (SEA) using an HEA.

6. When the Host Ethernet Adapters window appears again, press **OK**.

Important: If you make any changes in the HEA port properties, except for the Pending Port Group Multi-Core Scaling value, it will cause the link to go down and then back up.

If you change Pending Port Group Multi-Core Scaling value, this change will not take effect until the next system IPL.

Configuring a Integrated Virtual Ethernet adapter as an SEA

You can define an LHEA at the time of creating a partition (including a Virtual I/O Server partition), or dynamically, just like the other virtual adapters.

1. In the navigation pane, expand **Systems Management** → **Servers**, and select the system on which you desire to create a new partition.

2. In the Tasks pane (the lower right of the Hardware Management Console Workplace), expand **Configuration** → **Create Logical Partition**, and select **VIO Server** to start the Create Lpar Wizard.

3. In the Create Lpar Wizard, proceed through the Virtual I/O Server configuration steps up to **Logical Host Ethernet Adapters (LHEA)**.

4. In Logical Host Ethernet Adapters (LHEA) window, choose an HEA and select the physical port of the HEA to configure an LHEA, and then press the **Continue...** button (Figure A-5).
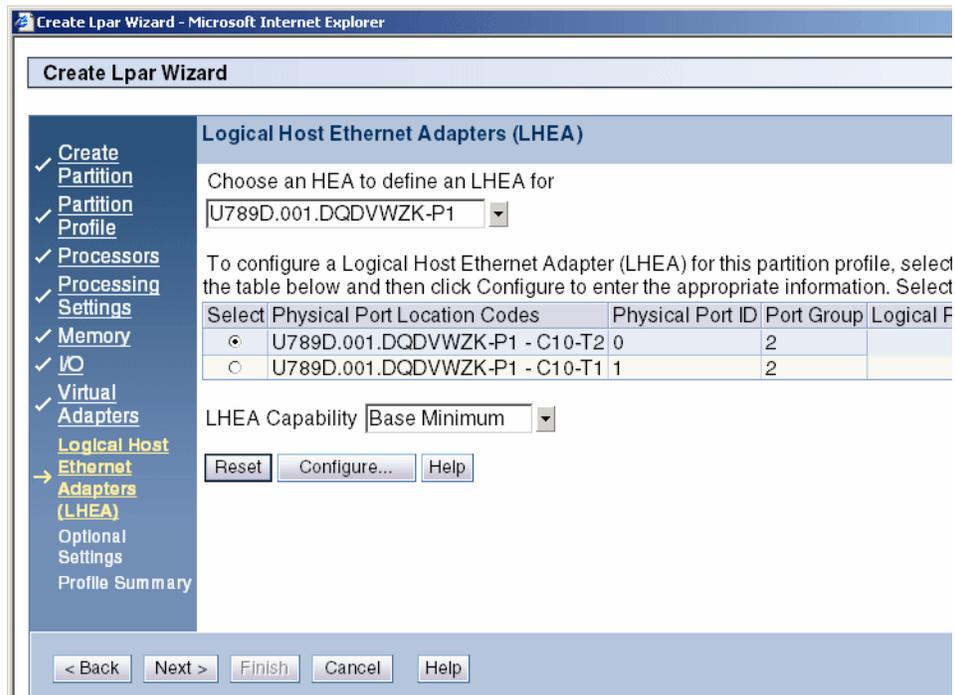
Figure A-5   Logical Host Ethernet Adapters (LHEA) window

5. Select a logical port in Logical Host Ethernet Adapter (LHEA) Configuration window, and press **OK**. You can optionally set VLAN IDs of the selected physical port (Figure A-6).
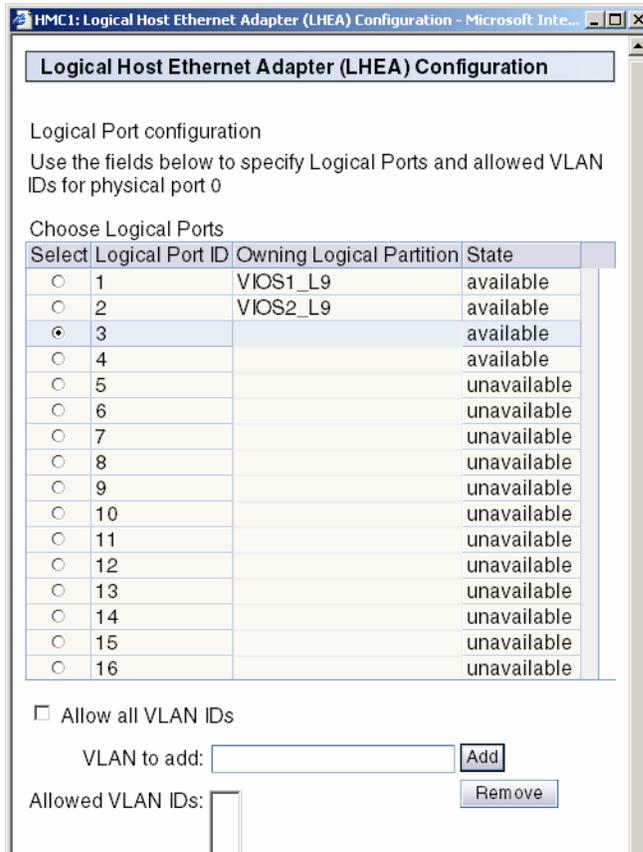
Figure A-6   Logical Host Ethernet Adapter (LHEA) Configuration window

6. When the Logical Host Ethernet Adapters (LHEA) window appears again, press **Next** and follow the steps.

> **Note:** The LHEA assigned to the partition is not shown in the Profile Summary window.

7. Install the Virtual I/O Server software on the partition.

8. Check if the logical HEA (LHEA) and the logical port are created on the activated Virtual I/O Server, as shown in Example A-1.

Example A-1   List of adapters on a Virtual I/O Server

```
$ lsdev -type adapter
name       status     description
ent0       Available  Logical Host Ethernet Port (lp-hea)
ent1       Available  Virtual I/O Ethernet Adapter (l-lan)
fcs0       Available  FC Adapter
lhea0      Available  Logical Host Ethernet Adapter (l-hea)
vasi0      Available  Virtual Asynchronous Services Interface (VASI)
vhost0     Available  Virtual SCSI Server Adapter
vsa0       Available  LPAR Virtual Serial Adapter
```

9. Create an SEA using the logical port of the HEA shown as lp-hea in Example A-2, treating it just as a normal Ethernet adapter on a Virtual I/O Server.

Example A-2   Creating a shared Ethernet Adapter

```
$ mkvdev -sea ent0 -vadapter ent1 -default ent1 -defaultid 1
$ lsdev -type adapter
name       status     description
ent0       Available  Logical Host Ethernet Port (lp-hea)
ent1       Available  Virtual I/O Ethernet Adapter (l-lan)
ent2       Available  Shared Ethernet Adapter
fcs0       Available  FC Adapter
lhea0      Available  Logical Host Ethernet Adapter (l-hea)
vasi0      Available  Virtual Asynchronous Services Interface (VASI)
vhost0     Available  Virtual SCSI Server Adapter
vsa0       Available  LPAR Virtual Serial Adapter
```

You can also add an LHEA to a running partition by using dynamic LPAR operation.

1. In the navigation pane, expand **Systems Management** → **Servers**, and select the system that has the partition to which you wish to add an LHEA.

2. In the Contents pane (the top right of the Hardware Management Console Workplace), select the partition to which you want to add an LHEA.

3. Click **View pop-up menu** and select **Dynamic Logical Partitioning** → **Host Ethernet** → **Add**.

4. In the Add Logical HEA Resources window, select the physical port of the HEA to configure an LHEA, and then press the **Continue...** button.

5. Select a logical port in Logical Host Ethernet Adapter (LHEA) Configuration window, and press **OK**. You can optionally set VLAN IDs of the selected physical port.

6. When the Add Logical HEA Resources window appears again, press **OK**.

7. Configure the LHEA on the partition by using the `cfgmgr` (AIX partition) or `cfgdev` (VIOS partition) command.

B

Error codes and logs

This Appendix lists System Reference Codes (SRCs) and operating system error logs that pertain to Live Partition Mobility and is divided into the following sections:

► "SRCs" on page 184

► "Error SRCs" on page 185

► "Operating system error logs" on page 186

SRCs

Table B-1 gives a list of SRCs that indicate the current state of a partition migration.

Table B-1   Progress SRCs

Code	Meaning
2004	Partition is performing the `drmgr` command. It is displayed on the source server while the partition is waiting to suspend and on the destination server until the `drmgr` processing has completed.
C2001020	Partition is source of an inactive migration.
C2001030	Partition is the target of an inactive migration.
C2001040	Partition is the target of an active migration.
C2001080	Partition processors are stopped.
C2001082	Partition processors are restarted.
C20010FF	Migration complete.
D200A250	Partition has requested to suspend as part of an active migration.
D200AFFF	Partition migration was canceled.

Error SRCs

Table B-2 gives a list of SRC codes that indicate problems with a partition migration.

Table B-2   SRC error codes

Code	Meaning
B2001130	Partition migration readiness check failed.
B2001131	Resume of LpQueues failed.
B2001132	Allocated LpEvents failed.
B2001133	Failed to lock the partition configuration of the partition.
B2001134	Failed to unlock the partition configuration of the partition.
B2001140	Processing of transferred data failed.
B2001141	Processing of transferred data failed.
B2001142	Processing of transferred data failed.
B2001143	Processing of transferred data failed.
B2001144	Failed to suspend virtual I/O for the partition.
B2001145	Failed to resume virtual I/O for the partition.
B2001151	Partition attempted a dump during migration.
B2002210	Data import failure.
B2002220	Data import failure.
B2008160	PFDS build failure.

Operating system error logs

Table B-3 gives a list of entries that may appear in the operating system error logs of the partitions involved in a partition migration. The first column of the table lists the label of the entry, and the second column lists which partition logs the entry.

Table B-3   Operating system error log entries

Error log entries	Location
CLIENT_FAILURE	Virtual I/O Server providing VSCSI services to mobile partition
MVR_FORCE_SUSPEND	Mover Service Partition Virtual I/O Server
MVR_MIG_COMPLETED	Mover Service Partition Virtual I/O Server
MVR_MIG_ABORTED	Mover Service Partition Virtual I/O Server
CLIENT_PMIG_STARTED	AIX 5L mobile partition
CLIENT_PMIG_DONE	AIX 5L mobile partition

Abbreviations and acronyms

ABI	Application Binary Interface	**CEC**	Central Electronics Complex
ACL	Access Control List	**CHRP**	Common Hardware Reference Platform
AFPA	Adaptive Fast Path Architecture	**CLI**	Command-Line Interface
AIO	Asynchronous I/O	**CLVM**	Concurrent LVM
AIX	Advanced Interactive Executive	**CPU**	Central Processing Unit
APAR	Authorized Program Analysis Report	**CRC**	Cyclic Redundancy Check
		CSM	Cluster Systems Management
API	Application Programming Interface	**CSV**	Comma Separated Values
ARP	Address Resolution Protocol	**CUoD**	Capacity Upgrade on Demand
ASMI	Advanced System Management Interface	**DCM**	Dual Chip Module
ATS	Advanced Technical Support	**DES**	Data Encryption Standard
BFF	Backup File Format	**DGD**	Dead Gateway Detection
BIND	Berkeley Internet Name Domain	**DHCP**	Dynamic Host Configuration Protocol
BIST	Built-In Self-Test	**DLPAR**	Dynamic LPAR
BLV	Boot Logical Volume	**DMA**	Direct Memory Access
BOOTP	Boot Protocol	**DNS**	Domain Naming System
BOS	Base Operating System	**DRM**	Dynamic Reconfiguration Manager
BSD	Berkeley Software Distribution	**DR**	Dynamic Reconfiguration
BSR	Barrier Synchronization Registers	**DVD**	Digital Versatile Disk
CA	Certificate Authority	**EC**	EtherChannel
CATE	Certified Advanced Technical Expert	**ECC**	Error Checking and Correcting
CD	Compact Disk	**EOF**	End of File
CDE	Common Desktop Environment	**EPOW**	Environmental and Power Warning
CD-R	CD Recordable	**ERRM**	Event Response resource manager
CD-ROM	Compact Disk-Read Only Memory	**ESS**	Enterprise Storage Server®
		F/C	Feature Code

FC	Fibre Channel	**JFS**	Journaled File System
FCAL	Fibre Channel Arbitrated Loop	**JIT**	Just in Time
FDX	Full Duplex	**L1**	Level 1
FLOP	Floating Point Operation	**L2**	Level 2
FRU	Field Replaceable Unit	**L3**	Level 3
FTP	File Transfer Protocol	**LA**	Link Aggregation
GDPS®	Geographically Dispersed Parallel Sysplex™	**LACP**	Link Aggregation Control Protocol
GID	Group ID	**LAN**	Local Area Network
GPFS™	General Parallel File System™	**LDAP**	Lightweight Directory Access Protocol
GUI	Graphical User Interface	**LED**	Light Emitting Diode
HACMP	High Availability Cluster Multiprocessing	**LHEA**	Logical Host Ethernet Adapter
		LMB	Logical Memory Block
HBA	Host Bus Adapters	**LPAR**	Logical Partition
HEA	Host Ethernet Adapter	**LPP**	Licensed Program Product
HMC	Hardware Management Console	**LUN**	Logical Unit Number
HPT	Hardware Page Table	**LV**	Logical Volume
HTML	Hypertext Markup Language	**LVCB**	Logical Volume Control Block
HTTP	Hypertext Transfer Protocol	**LVM**	Logical Volume Manager
Hz	Hertz	**MAC**	Media Access Control
I/O	Input/Output	**Mbps**	Megabits Per Second
IBM	International Business Machines	**MBps**	Megabytes Per Second
		MCM	Multichip Module
ID	Identification	**ML**	Maintenance Level
IDE	Integrated Device Electronics	**MP**	Multiprocessor
IEEE	Institute of Electrical and Electronics Engineers	**MPIO**	Multipath I/O
		MSP	Mover Service Partition
IP	Internetwork Protocol	**MTU**	Maximum Transmission Unit
IPAT	IP Address Takeover	**NFS**	Network File System
IPL	Initial Program Load	**NIB**	Network Interface Backup
IPMP	IP Multipathing	**NIM**	Network Installation Management
ISV	Independent Software Vendor		
ITSO	International Technical Support Organization	**NIMOL**	NIM on Linux
		NTP	Network Time Protocol
IVM	Integrated Virtualization Manager	**NVRAM**	Non-Volatile Random Access Memory

ODM	Object Data Manager	**RISC**	Reduced Instruction-Set Computer
OSPF	Open Shortest Path First	**RMC**	Resource Monitoring and Control
PCI	Peripheral Component Interconnect	**RPC**	Remote Procedure Call
PIC	Pool Idle Count	**RPL**	Remote Program Loader
PID	Process ID	**RPM**	Red Hat Package Manager
PKI	Public Key Infrastructure	**RSA**	Rivet, Shamir, Adelman
PLM	Partition Load Manager	**RSCT**	Reliable Scalable Cluster Technology
PMAPI	Performance Monitor API	**RSH**	Remote Shell
PMP	Project Management Professional	**SAN**	Storage Area Network
POST	Power-On Self-test	**SCSI**	Small Computer System Interface
POWER	Performance Optimization with Enhanced Risc (Architecture)	**SDD**	Subsystem Device Driver
		SEA	Shared Ethernet Adapter
PPC	Physical Processor Consumption	**SIMD**	Single Instruction Multiple Data
PPFC	Physical Processor Fraction Consumed	**SMIT**	System Management Interface Tool
PTF	Program Temporary Fix	**SMP**	Symmetric Multiprocessor
PTX®	Performance Toolbox	**SMS**	System Management Services
PURR	Processor Utilization Resource Register	**SMT**	Simultaneous Muli-threading
PV	Physical Volume	**SP**	Service Processor
PVID	Physical Volume Identifier	**SPOT**	Shared Product Object Tree
PVID	Port Virtual LAN Identifier	**SRC**	System Resource Controller
QoS	Quality of Service	**SRN**	Service Request Number
RAID	Redundant Array of Independent Disks	**SSA**	Serial Storage Architecture
RAM	Random Access Memory	**SSH**	Secure Shell
RAS	Reliability, Availability, and Serviceability	**SSL**	Secure Socket Layer
		SUID	Set User ID
RCP	Remote Copy	**SVC**	SAN Virtualization Controller
RDAC	Redundant Disk Array Controller	**TCP/IP**	Transmission Control Protocol/Internet Protocol
RIO	Remote I/O	**TSA**	Tivoli® System Automation
RIP	Routing Information Protocol	**UDF**	Universal Disk Format
		UDID	Universal Disk Identification

VASI	Virtual asynchronous services interface
VIPA	Virtual IP Address
VG	Volume Group
VGDA	Volume Group Descriptor Area
VGSA	Volume Group Status Area
VLAN	Virtual Local Area Network
VP	Virtual Processor
VPD	Vital Product Data
VPN	Virtual Private Network
VRRP	Virtual Router Redundancy Protocol
VSD	Virtual Shared Disk
WLM	Workload Manager

Related publications

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

IBM Redbooks

For information about ordering these publications, see "How to get IBM Redbooks" on page 194. Note that some of the documents referenced here may be available in softcopy only.

► *AIX 5L Differences Guide Version 5.3 Edition*, SG24-7463

► *AIX 5L Practical Performance Tools and Tuning Guide*, SG24-6478

► *Effective System Management Using the IBM Hardware Management Console for pSeries*, SG24-7038

► *IBM System p Advanced POWER Virtualization Best Practices*, REDP-4194

► *Implementing High Availability Cluster Multi-Processing (HACMP) Cookbook*, SG24-6769

► *Introduction to pSeries Provisioning*, SG24-6389

► *Linux Applications on pSeries*, SG24-6033

► *Managing AIX Server Farms*, SG24-6606

► *NIM from A to Z in AIX 5L*, SG24-7296

► *Partitioning Implementations for IBM eServer p5 Servers*, SG24-7039

► *A Practical Guide for Resource Monitoring and Control (RMC)*, SG24-6615

► *Integrated Virtualization Manager on IBM System p5*, REDP-4061

Other publications

These publications are also relevant as further information sources:

► The following types of documentation are located through the Internet at the
  following URL:

  http://www.ibm.com/servers/eserver/pseries/library

  – User guides

  – System management guides

  – Application programmer guides

  – All commands reference volumes

  – Files reference

  – Technical reference volumes used by application programmers

► Detailed documentation about the Advanced POWER Virtualization feature
  and the Virtual I/O Server

  https://www14.software.ibm.com/webapp/set2/sas/f/vios/documentation/
  home.html

► *AIX 5L V5.3 Partition Load Manager Guide and Reference,* SC23-4883

► *Linux for pSeries installation and administration (SLES 9)*, found at:

  http://www-128.ibm.com/developerworks/linux/library/l-pow-pinstall/

► *Linux virtualization on POWER5: A hands-on setup guide*, found at:

  http://www-128.ibm.com/developerworks/edu/l-dw-linux-pow-virutal.htm
  l

► *POWER5 Virtualization: How to set up the SUSE Linux Virtual I/O Server*,
  found at:

  http://www-128.ibm.com/developerworks/eserver/library/es-susevio/

Online resources

These Web sites and URLs are also relevant as further information sources:

► AIX and Linux on POWER community

  http://www.ibm.com/systems/p/community/

► Capacity on Demand

  http://www.ibm.com/systems/p/cod/

- ► IBM Advanced POWER Virtualization on IBM System p Web page

  http://www.ibm.com/systems/p/apv/

- ► IBM eServer pSeries and AIX Information Center

  http://publib16.boulder.ibm.com/pseries/index.htm

- ► IBM System Planning Tool

  http://www.ibm.com/servers/eserver/support/tools/systemplanningtool/

- ► IBM Systems Hardware Information Center

  http://publib.boulder.ibm.com/infocenter/eserver/v1r3s/index.jsp

- ► IBM Systems Workload Estimator

  http://www-912.ibm.com/supporthome.nsf/document/16533356

- ► Latest *Multipath Subsystem Device Driver User's Guide*

  http://www.ibm.com/support/docview.wss?rs=540&context=ST52G7&uid=ssg
  1S7000303

- ► Novell SUSE Linux Enterprise Server information

  http://www.novell.com/products/server/index.html

- ► SCSI T10 Technical Committee

  http://www.t10.org

- ► SDDPCM software download page

  http://www.ibm.com/support/docview.wss?uid=ssg1S4000201

- ► SDD software download page

  http://www.ibm.com/support/docview.wss?rs=540&context=ST52G7&dc=D430
  &uid=ssg1S4000065&loc=en_US&cs=utf-8&lang=en

- ► Service and productivity tools for Linux on POWER

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

- ► Silicon-on-insulator (SOI) technology

  http://www.ibm.com/chips/technology/technologies/soi/

- ► VIOS supported environment

  http://techsupport.services.ibm.com/server/vios/documentation/datash
  eet.html

- ► Virtual I/O Server documentation

  http://techsupport.services.ibm.com/server/vios/documentation/home.h
  tml

- ► Virtual I/O Server home page

  http://techsupport.services.ibm.com/server/vios/home.html
- ► Virtual I/O Server home page (alternate)

  http://www14.software.ibm.com/webapp/set2/sas/f/vios/home.html
- ► Virtual I/O Server supported hardware

  http://www14.software.ibm.com/webapp/set2/sas/f/vios/documentation/d
  atasheet.html
- ► Virtual I/O Server Support Page

  http://techsupport.services.ibm.com/server/vios/download/home.html

How to get IBM Redbooks

You can search for, view, or download Redbooks, IBM Redpapers, Technotes, draft publications and Additional materials, as well as order hardcopy Redbooks, at this Web site:

ibm.com/redbooks

Help from IBM

IBM Support and downloads

ibm.com/support

IBM Global Services

ibm.com/services

Index

Symbols

139, 141–142, 144–149

A

accounting, AIX   39
active migration   4, 29
  capability   21, 31
  compatibility   21, 31
  completion   36
  concurrent   32
  configuration checks   31
  definition   4
  dirty page   35
  entitlements   23
  example   8
  HMC   12
  memory modification   35
  migratability   23
  migration phase   33
  mover service partition   12
  MSP selection   37
  multiple concurrent migrations   121
  preparation   30
  prerequisites   23
  reactivation   36
  recovery   166
  requirements   122
  shared Ethernet adapter   7
  state   29
  stopping   38
  validation   164, 166
  VIOS selection   37
  workflow   5, 12
active profile   26
adapters
  dedicated   23, 30–31
  physical   23, 27, 31
  virtual   27
advanced accounting, AIX   39
Advanced POWER Virtualization   3
AIX 5L   26, 32, 39, 44
  kernel extensions   36, 39
  Live Workload Partitions   15

alternate error log   32
applications
  check-migrate   32
  migration capability   32
  prepare for migration   35
  reconfiguration notification   36
ARP   36
ASMI   47
attribute
  mover service partition   87
  reserve_policy   86
  time reference   20, 87
availability checks   32
availability requirement   4

B

barrier synchronization register   21, 32, 65
basic environments   84
battery   47
battery power   22
bootlist command   132
bootlist, AIX command   132
bosboot, command   132
BSR   21, 32

C

capability   26
  active migration   31
  operating system   32
Capacity on Demand   50, 52
cfgmgr, command   130, 133
changes
  non-reversible   29
  reversible   29
  rollback   29, 38
chdev command   74
chdev, VIOS command   126
check-migrate request   32
chhwres command   149
chsyscfg command   148
chvg, AIX command   131
CLI   37–151
command-line

195

IBM System p Live Partition Mobility

Redbooks

IBM System p
Live Partition Mobility

Redbooks

Explore the POWER6 processor-based live partition mobility

Perform examples of how to move running partitions between servers

Learn how to automate partition migration

Live Partition Mobility allows you to migrate running AIX and Linux partitions and their hosted applications from one physical server to another without disruption to either the infrastructure services or the users. The migration transfers the entire partition state, including processor context, memory, attached virtual devices, and connected users.

Live Partition Mobility helps you meet increasingly stringent service-level agreements (SLAs) because it allows you to proactively move running partitions and applications from systems requiring intervention or heavily loaded servers on to other systems.

Live Partition Mobility can be automated and incorporated in to system management tools and scripts. Support for multiple concurrent migrations allows you to liberate systems very quickly. For single partition, point-in-time migrations, the HMC offers an easy-to-use migration wizard.