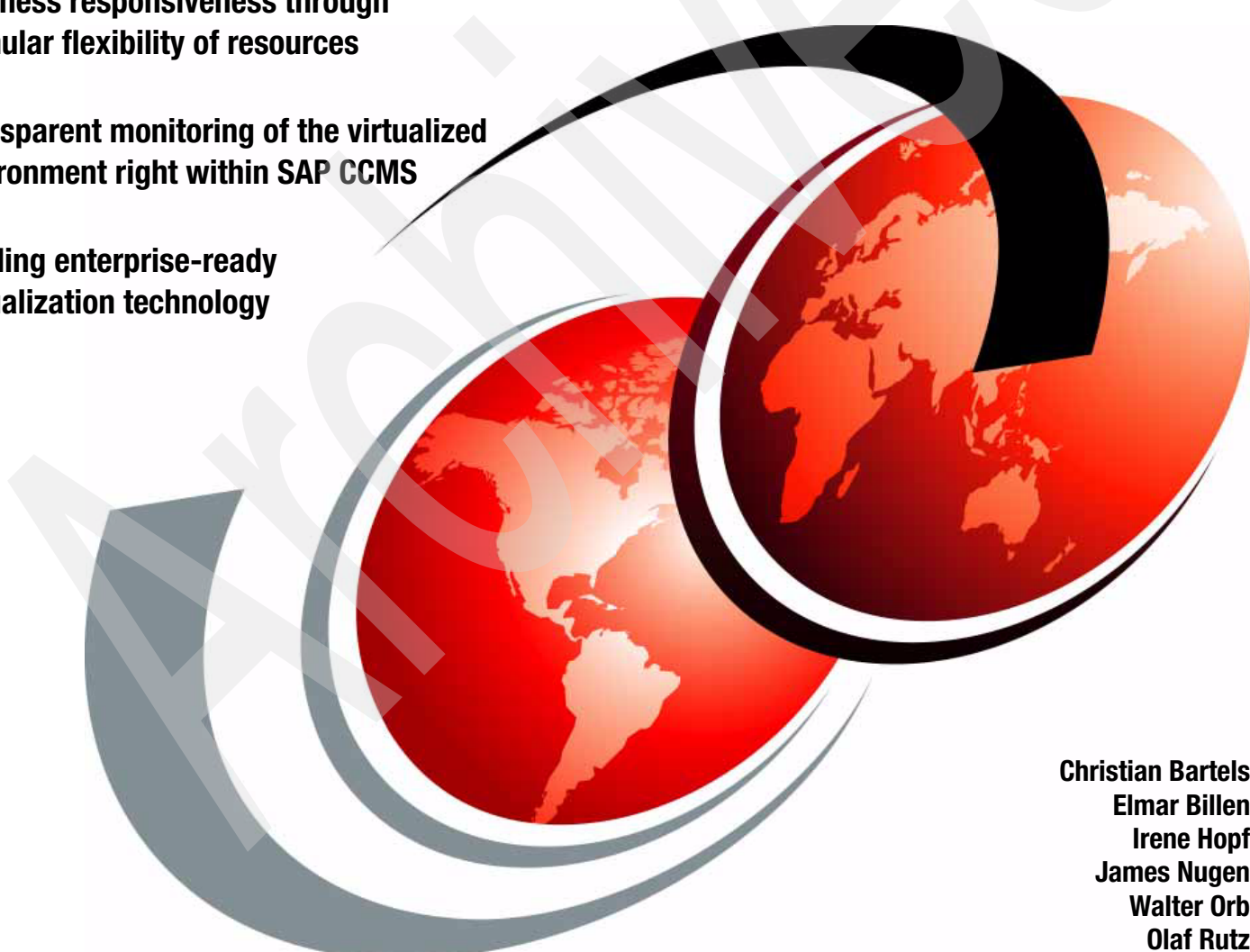


# SAP Applications on IBM PowerVM

**Business responsiveness through granular flexibility of resources**

**Transparent monitoring of the virtualized environment right within SAP CCMS**

**Leading enterprise-ready virtualization technology**



**Christian Bartels  
Elmar Billen  
Irene Hopf  
James Nugen  
Walter Orb  
Olaf Rutz**

**Redbooks**





International Technical Support Organization

**SAP Applications on IBM PowerVM**

October 2011

Archived

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

Archived

**Second Edition (October 2011)**

This edition applies to AIX 7.1, AIX 6.1, AIX 5.3, IBM i 7.1, IBM i 6.1, IBM i 5.4, SUSE Linux Enterprise Server 11 (SLES 11), SUSE Linux Enterprise Server 10 (SLES 10), SUSE Linux Enterprise Server 9 (SLES 9), Red Hat Enterprise Linux 5 (RHEL 5), and Red Hat Enterprise Linux 4 (RHEL 4)

© Copyright International Business Machines Corporation 2008, 2011. All rights reserved.

Note to U.S. Government Users Restricted Rights -- Use, duplication or disclosure restricted by GSA ADP Schedule Contract with IBM Corp.

# Contents

<b>Notices</b> .....	vii
Trademarks .....	viii
<b>Preface</b> .....	ix
The teams who wrote this book .....	x
Now you can become a published author, too! .....	xii
Comments welcome .....	xiii
Stay connected to IBM Redbooks .....	xiii
<b>Summary of changes</b> .....	xv
October 2011, Second Edition .....	xv
<b>Chapter 1. From a non-virtualized to a virtualized infrastructure</b> .....	1
1.1 Motivation .....	1
1.1.1 Evolution trends in the SAP landscape .....	1
1.1.2 Motivation for server virtualization .....	4
1.1.3 Beyond the server .....	4
1.2 Examples of virtualization advantages .....	4
1.2.1 Landscape consolidation .....	4
1.2.2 Load shifting .....	5
1.2.3 Processing chain .....	7
1.2.4 Concurrent load prioritization .....	8
<b>Chapter 2. PowerVM virtualization technologies</b> .....	11
2.1 Hypervisor .....	13
2.2 Hardware Management Console .....	14
2.3 Integrated Virtualization Manager .....	14
2.4 Systems Director Management Console .....	15
2.5 Systems Director VMControl .....	15
2.6 Dedicated LPARs .....	16
2.7 Live Partition Mobility .....	16
2.8 Dynamic LPAR .....	16
2.9 Micropartitioning and Shared Processor LPARs .....	17
2.10 Shared Dedicated Capacity .....	18
2.11 Multiple Shared-Processor Pools .....	18
2.12 Virtual I/O Server .....	19
2.13 Partition Suspend and Resume .....	19
2.14 N Port ID Virtualization .....	20
2.15 Virtual Tape .....	20
2.16 Virtual SCSI .....	20
2.17 Virtual Ethernet .....	21
2.18 Shared Ethernet Adapter .....	22
2.19 Integrated Virtual Ethernet .....	23
2.20 Active Memory Sharing .....	23
2.21 Active Memory Expansion .....	24
2.22 Workload Partitions .....	25
2.23 Workload Partition Manager .....	25
2.24 Live Application Mobility .....	25
2.25 Simultaneous Multithreading .....	25

2.26 IBM i subsystems . . . . .	26
<b>Chapter 3. Best practice implementation example at a customer site. . . . .</b>	<b>27</b>
<b>Chapter 4. Hands-on management tasks . . . . .</b>	<b>29</b>
4.1 Ease of moving from dedicated to shared. . . . .	29
4.2 Micropartition design option . . . . .	32
4.3 Assigning implicit capping to LPARs. . . . .	33
4.4 Scheduling dynamic LPAR operations . . . . .	34
4.5 Managing multiple shared processor pools. . . . .	36
4.6 Enable Pool Utilization Authority . . . . .	38
4.7 Activate and modify memory compression . . . . .	38
4.8 Simultaneous Multithreading and compatibility modes . . . . .	40
<b>Chapter 5. Virtual I/O Server . . . . .</b>	<b>43</b>
5.1 Motivation . . . . .	44
5.2 Virtual I/O Server: basic usage types . . . . .	44
5.2.1 Using virtual SCSI. . . . .	45
5.2.2 High availability for vSCSI. . . . .	45
5.2.3 Using the virtual network. . . . .	46
5.2.4 High availability for virtual networks . . . . .	47
5.3 Setting up a VIOS partition . . . . .	48
5.3.1 Defining the VIOS LPAR. . . . .	48
5.3.2 Installing the VIOS . . . . .	48
5.3.3 Creating virtual SCSI Server Adapters . . . . .	48
5.3.4 Gathering information about existing virtual adapters. . . . .	50
5.3.5 Connecting a client LPAR to a virtual SCSI server adapter . . . . .	50
5.3.6 Creating virtual Ethernet adapters . . . . .	52
5.3.7 Connecting a client LPAR to the virtual network. . . . .	52
5.3.8 TCP/IP address for the VIOS . . . . .	53
5.3.9 N_Port ID Virtualization. . . . .	54
5.3.10 VIOS backup. . . . .	54
5.4 VIO Server monitoring. . . . .	54
<b>Chapter 6. IBM PowerVM Live Partition Mobility . . . . .</b>	<b>63</b>
6.1 Introduction . . . . .	63
6.2 Prerequisites . . . . .	65
6.3 Performance . . . . .	66
6.4 SAP notes . . . . .	67
6.5 SAP license keys . . . . .	67
6.6 Flexible License mechanism. . . . .	67
6.7 Demo of Live Partition Mobility with running SAP systems. . . . .	68
6.8 POWER6 to POWER7 migrations . . . . .	68
<b>Chapter 7. Workload partitions . . . . .</b>	<b>69</b>
7.1 Characteristics of WPARs. . . . .	70
7.2 Types of WPARs. . . . .	70
7.2.1 System WPARs . . . . .	70
7.2.2 Application WPARs. . . . .	71
7.2.3 Versioned WPARs . . . . .	71
7.3 Support of SAP applications running in WPARs. . . . .	71
7.3.1 Supported scenarios . . . . .	72
7.4 Sample installation of SAP NetWeaver 7.0 in a Shared System WPAR. . . . .	73
7.5 Live Application Mobility . . . . .	74

<b>Chapter 8. SAP system setup for virtualization</b> .....	75
8.1 SAP Adaptive Computing .....	76
8.1.1 Overview .....	76
8.1.2 Adaptive Computing compliance test .....	77
8.1.3 Technical Implementation .....	77
8.2 SAP instance profile changes .....	78
8.2.1 Dynamic work processes .....	78
8.2.2 Extended memory tuning recommendations for AIX .....	78
8.2.3 Extended memory tuning recommendations for IBM i .....	79
8.3 Virtual Memory tuning in AIX for SAP systems .....	80
8.4 Main storage pools, work process priorities, and workload capping on IBM i .....	82
8.4.1 Separation of main storage pools .....	82
8.4.2 Work process priorities and workload capping .....	83
8.5 Active Memory Expansion for SAP systems .....	84
8.6 Processor utilization metrics .....	85
8.6.1 Introduction .....	85
8.6.2 Test case .....	86
8.6.3 Processor time measurement and SMT .....	87
8.6.4 PURR-based metrics in POWER7 .....	91
8.6.5 Processor time in SAP applications .....	92
8.6.6 Side effects of PURR based metrics in Shared Pool environments .....	94
<b>Chapter 9. Monitoring</b> .....	95
9.1 Performance Monitoring in AIX .....	96
9.1.1 The vmstat monitoring tool .....	96
9.1.2 The sar monitoring tool .....	97
9.1.3 The mpstat tool .....	99
9.1.4 The lparstat tool .....	100
9.1.5 The topas tool .....	103
9.1.6 The nmon tool .....	106
9.2 Performance monitoring on IBM i .....	108
9.2.1 Ad-hoc Performance Tools on IBM i .....	108
9.2.2 Performance data collection tools on IBM i .....	111
9.2.3 Performance Data Analysis Tools on IBM i .....	113
9.2.4 Enhanced tools for root cause analysis .....	116
9.3 Other monitoring tools .....	117
9.3.1 LPAR2rrd .....	117
9.3.2 Ganglia .....	121
9.3.3 Tivoli Monitoring .....	126
9.4 Monitoring of virtualized SAP systems .....	127
9.4.1 Motivation and challenges .....	127
9.4.2 SAP CCMS operating system monitor .....	128
9.4.3 SAP Monitoring Infrastructure .....	136
9.5 SAP EarlyWatch and SAP GoingLive check services .....	156
9.6 IBM Insight for SAP .....	157
9.6.1 Installing and configuring IBM Insight for SAP .....	158
9.6.2 Data collection .....	159
9.6.3 Report Generation Service .....	159
<b>Chapter 10. Support statements by IBM and SAP</b> .....	161
10.1 SAP general support statement .....	162
10.2 AIX .....	162
10.3 IBM i .....	162
10.4 Linux .....	162

<b>Related publications</b> .....	165
IBM Redbooks .....	165
Online resources .....	165
Documentation about Live Partition Mobility .....	167
White papers about Live Partition Mobility .....	167
Other resources .....	168
How to get IBM Redbooks .....	168
Help from IBM .....	168
<b>Index</b> .....	169



# 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.

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

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


## COPYRIGHT LICENSE:

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

# Trademarks

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

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

Active Memory™	Power Systems™	pSeries®
AIX 5L™	POWER4™	Redbooks®
AIX®	POWER5™	Redbooks (logo)  ®
DB2®	POWER6+™	System i®
IBM®	POWER6®	System p®
iSeries®	POWER7™	System x®
Micro-Partitioning™	PowerHA™	Systems Director VMControl™
OS/400®	PowerVM™	Tivoli®
POWER Hypervisor™	POWER®	Workload Partitions Manager™

The following terms are trademarks of other companies:

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

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

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

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

Novell, SUSE, the Novell logo, and the N logo are registered trademarks of Novell, Inc. in the United States and other countries.

Oracle, JD Edwards, PeopleSoft, Siebel, and TopLink are registered trademarks of Oracle Corporation and/or its affiliates.

ABAP, SAP EarlyWatch, SAP GoingLive, SAP NetWeaver, SAP, and SAP logos are trademarks or registered trademarks of SAP AG in Germany and in several other countries.

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

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

The term virtualization leads to a wealth of definitions. A generic, rather popular scientific description is "In computing, virtualization is a broad term that refers to the abstraction of computer resources."

IBM® invented the virtualization technology starting in the 1960s on the mainframe, and the functionalities evolved and were ported to other platforms and improved the reliability, availability, and serviceability (RAS) features. With virtualization, you achieve better asset utilization, reduced operating costs, and faster responsiveness to changing business demands.

Every technology vendor in the SAP ecosystem understands virtualization as slightly different capabilities on different levels (storage and server hardware, processor, memory, I/O resources or the application, and so on). It is important to understand exactly what functionality is offered and how it supports the client's business requirements.

SAP covers the subject in their SAP Community Network (SDN) web page at:

<https://www.sdn.sap.com/irj/sdn/virtualization>

In this IBM Redbooks® publication we focus on server virtualization technologies in the IBM Power Systems™ hardware, AIX®, IBM i, and Linux space and what they mean specifically for SAP applications running on this platform.

We do not repeat information that is already available from other sources unless it is directly related to the IT infrastructure for the SAP solution; therefore, we included a lot of links and references for further information about virtualization in general.

SAP clients can leverage the technology that the IBM Power Systems platform offers. In this book, we describe the technologies and functions, what they mean, and how they apply to the SAP system landscape.

SAP currently has no limitations, in general, supporting their products that are running on PowerVM™ technology. If there are specific limitations, we mention them in the individual chapters and compile them in the list of links in "Related publications" on page 165.

This publication contains the following content with regard to virtualization and SAP applications on IBM PowerVM:

- ▶ From a non-virtualized to a virtualized infrastructure
- ▶ PowerVM virtualization technologies
- ▶ Best practice implementation example at a customer
- ▶ Hands-on management tasks
- ▶ Virtual I/O Server
- ▶ IBM PowerVM Live Partition Mobility
- ▶ Workload partitions
- ▶ SAP system setup for virtualization
- ▶ Monitoring
- ▶ Support statements by SAP and IBM

## The teams who wrote this book

The first edition of this book was produced by two teams of specialists from around the world working in Walldorf Germany and at the International Technical Support Organization, Austin Texas Center in 2008.

The authors of the first edition are:

The Austin, Texas team:

- ▶ Niels Abildskov,
- ▶ Carsten Dieterle,
- ▶ Wilhelm Koestinger,
- ▶ Niikolaj Lyngfeldt
- ▶ Mark Smith

The Walldorf, Germany team:

- ▶ Elmar Billen,
- ▶ Carol Davis,
- ▶ Katharina Probst,
- ▶ Irene Hopf,
- ▶ Jim Nugen,
- ▶ Olaf Rutz
- ▶ Andreas Steiner

In the summer of 2011 a team in Walldorf, Germany produced the update to this book:

**Christian Bartels** has a degree in Electrical Engineering from the Technical University Braunschweig, Germany. He joined IBM in 1990 and has been working in OS/400® Software Support in Mainz, Germany. In 1997 he moved to the SAP Regional Support center in Walldorf, Germany in order to support clients running SAP R/3 on OS/400 and to educate the SAP support team. Between 1999 and 2001, Christian worked at IBM in Rochester, MN in the ISV enablement team with a special focus on the database interface for SAP on OS/400 and performance analysis. Since 2001, he is working on the SAP on IBM i porting team. His current emphasis is on performance monitoring.

**Elmar Billen** is a Senior SAP Basis Consultant certified by SAP and Oracle. He has worked with SAP software since 1995 and has strong experience in using SAP on Linux and UNIX. He implemented several large SAP solutions and gained experience with system management, high availability, backup and recovery, migrations, and release upgrades. Since 2004 he has worked on the IBM Power® Linux team at SAP and participated in porting the SAP solutions to Linux. He has a degree in Physics from the University of Münster.

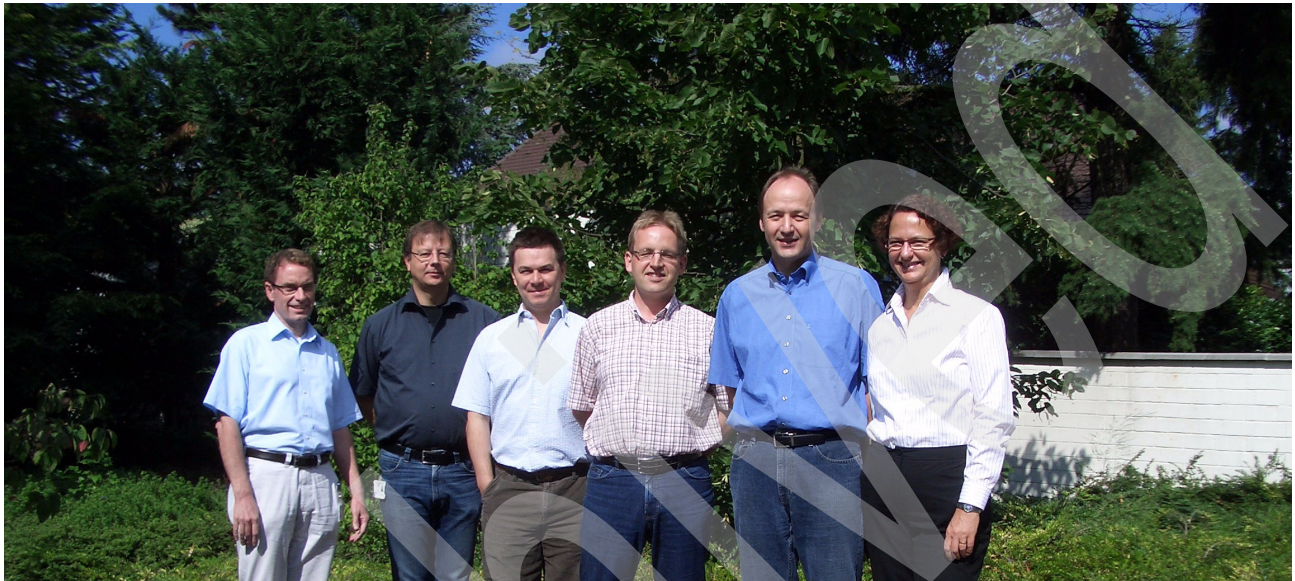
**Irene Hopf** is a Senior Architect in the IBM SAP International Competence Center (ISICC) in Walldorf, Germany. She is responsible for integrated IBM hardware and software solutions for SAP clients with the focus on flexible operations and optimal performance in a dynamic, cloud-ready infrastructure. Irene has more than 15 years of SAP experience and is a certified SAP Technology Consultant. She holds degrees in Material Science Engineering and Information Technology.

**James Nugen** has been a member of the AIX porting team at SAP since 1995. He is an IBM Certified Advanced Technical Expert for pSeries® and AIX 5L™. Jim has a Bachelor's degree in Electrical Engineering from Princeton University and an MS from the University of Reading.

**Walter Orb** is a technical consultant working at the IBM SAP International Competence Center in Walldorf, Germany. He has more than fifteen years experience with SAP on AIX and

Power Systems, with a major focus on system performance, benchmarks, and large-scale system tests. Walter holds a Master's degree in Mathematics and Economics from the University of Mainz, Germany.

**Olaf Rutz** studied at the Technical University in Darmstadt and has a Master's degree in Computer Science. He joined IBM in 1995 and since then worked in several SAP-related projects at SAP. During this time, he gained several years of experience in porting SAP applications to IBM mainframe and Power Systems with AIX and IBM i. In his current role, he is the Team Lead for the SAP on AIX team in St. Leon-Rot.



*Figure 0-1 From left: Christian Bartels, Olaf Rutz, James Nugen, Elmar Billen, Walter Orb and Irene Hopf*

Special thanks to Dr. Georg Leffers from SAP who provided valuable contributions to the update project and Scott Vetter, PMP at IBM Austin, who was our project manager, remotely.

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

- ▶ Ole Gadegaard  
IBM Denmark
- ▶ Jim Rymarczyk  
IBM Fellow
- ▶ Fabienne Lepetit  
Thomas Mariën  
IBM France
- ▶ Jochen Breh  
Egon Irmmler  
Matthias Koechl  
Martin Mezger  
Joachim Stumpf  
IBM Germany
- ▶ Saravanan Devendra  
IBM India
- ▶ Nigel Griffith  
IBM UK

- ▶ John Bailey  
Brad Cobb  
Katherine Dugan  
Jim Grecni  
Vinit Jain  
Greg Mussich  
Edward Shvartsman  
Michell Spooner  
Randy Swanberg  
IBM USA
- ▶ KaTrina Love Abram, ITSO Technical Editor  
Alfred Schwab, ITSO Technical Editor  
IBM USA

## Now you can become a published author, too!

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

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

[ibm.com/redbooks/residencies.html](http://ibm.com/redbooks/residencies.html)

## Comments welcome

Your comments are important to us!

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

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

[ibm.com/redbooks](http://ibm.com/redbooks)

- ▶ Send your comments in an email to:

[redbooks@us.ibm.com](mailto:redbooks@us.ibm.com)

- ▶ Mail your comments to:

IBM Corporation, International Technical Support Organization  
Dept. HYTD Mail Station P099

## Stay connected to IBM Redbooks

- ▶ Find us on Facebook:

<http://www.facebook.com/IBMRedbooks>

- ▶ Follow us on Twitter:

<http://twitter.com/ibmredbooks>

- ▶ Look for us on LinkedIn:

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

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

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

- ▶ Stay current on recent Redbooks publications with RSS Feeds:

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

Archived



# Summary of changes

This section describes the technical changes made in this edition of the book and in previous editions. This edition might also include minor corrections and editorial changes that are not identified.

Summary of Changes  
for SG24-7564-01  
for SAP Applications on IBM PowerVM  
as created or updated on November 1, 2011.

## October 2011, Second Edition

This revision reflects the addition, deletion, or modification of new and changed information described below.

### New information

- ▶ Throughout the book, IBM i related content was added to all relevant chapters.
- ▶ Description of new functionalities such as Systems Director Management Console, Active Memory Expansion, Active Memory Sharing, N Port ID Virtualization, Suspend Resume, Thin Provisioning and Simultaneous Multithreading 4 in 2.25, “Simultaneous Multithreading” on page 25.
- ▶ How to use Active Memory Expansion for SAP landscapes and how to set Pool Utilization Authority in Chapter 4, “Hands-on management tasks” on page 29.
- ▶ N Port ID Virtualization and VIOS Backup Restore in Chapter 5, “Virtual I/O Server” on page 43.
- ▶ Performance results and compatibility modes as an addition to Chapter 6, “IBM PowerVM Live Partition Mobility” on page 63.
- ▶ Explanation about CPU utilization metrics and the meaning of capacity consumed versus busy in 8.6, “Processor utilization metrics” on page 85 including comments about the usage of SMT-4 on Power7 based systems.

### Changed information

- ▶ Update and synchronize the feature function list in Table 2-1 on page 12.
- ▶ Minor adjustments in Chapter 7, “Workload partitions” on page 69 and added the topic “versioned WPARs”.
- ▶ Update to the integration between SAP Adaptive Computing and IBM Hardware Management Console or IBM Systems Director in 8.1, “SAP Adaptive Computing” on page 76.
- ▶ Major rework about monitoring of SAP landscapes on PowerVM:
  - Monitoring of Active Memory Expansion (option -c) in 9.1, “Performance Monitoring in AIX” on page 96
  - Usage of mpstat, lparstat, topas and topasrec in SAP landscapes 9.1, “Performance Monitoring in AIX” on page 96
  - Performance tools for IBM i in 9.2, “Performance monitoring on IBM i” on page 108

- Monitoring of VIOS with saposcol in SAP CCMS in 9.4, “Monitoring of virtualized SAP systems” on page 127
- ▶ Updated list of further resources and SAP Notes in “Related publications” on page 165

Archived

# From a non-virtualized to a virtualized infrastructure

In this chapter, we focus on:

- ▶ Motivation to move from a non-virtualized to a virtualized infrastructure:
  - Evolution in the SAP landscape
  - Motivation for server virtualization
- ▶ Examples of virtualization advantages:
  - Landscape consolidation
  - Load shifting
  - Processing chain
  - Concurrent load prioritization

## 1.1 Motivation

In this section, we focus on the evolutionary trends in the SAP landscape and the motivation for server virtualization.

### 1.1.1 Evolution trends in the SAP landscape

The classic architecture of an SAP system is a client/server design. The building blocks of a traditional SAP system comprise a database and a number of application servers. One of these application servers (apps svr), the central instance or CI, contains resources that are unique to the SAP system: the object locking mechanism (enq\_server) and the system communications mechanism (msg\_server). You can construct the system in either 2-tier (all components on a single server) or 3-tier (application servers distributed over multiple servers), as shown in Figure 1-1 on page 2 where the presentation layer (front-end users) counts as one separate tier.

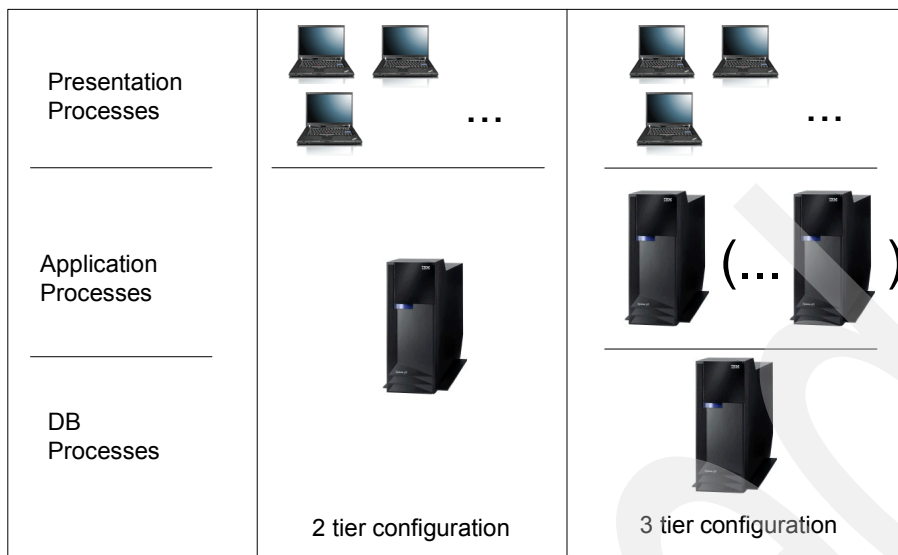


Figure 1-1 2-tier or 3-tier configuration

In a 2-tier configuration, we see a better resource utilization and the overall administration is easier. The 3-tier system architecture necessarily increases the number of systems that must be maintained and the number of operating system (OS) images. Nevertheless, a 3-tier can offer benefit in a number of ways: the ability of scale-out and in reducing the exposure in the case of a single-server failure and the granularity for a fail-over scenario.

You must also consider that an SAP production system never comes alone. In a normal SAP landscape, each production system represents a string of at least three systems: development, quality assurance, and production, as shown in Figure 1-2.

The development system is often much smaller than the production system; however, the test or quality assurance system is used to simulate production and, at least during major tests, must have the full capacity of the production system.

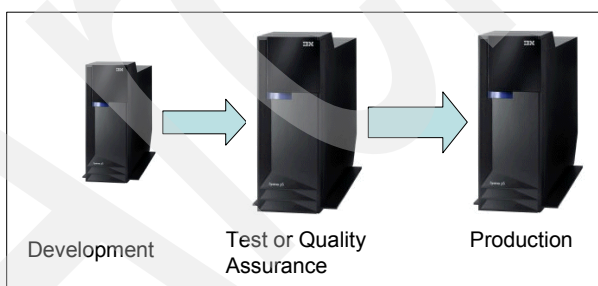


Figure 1-2 SAP system landscape

Up to now, we only looked at the traditional landscape for a single system. An SAP landscape normally contains a number of core systems of this kind that cover different application areas, depending on the industry, which includes, but is not limited to: Enterprise Resource Planning (ERP/ECC<sup>1</sup>), Supply Chain Management (SCM), Customer Relationship Management (CRM), Supplier Relationship Management (SRM), Business Intelligence (BI), and systems that are used for supporting functionality, such as the SAP Solution Manager. These systems are joined by functionality for interoperability, such as Process Integration (PI) for inter-system communication, and the Enterprise Portal (EP) for front-end communication, which each

<sup>1</sup> ECC = ERP Central Component

require development, test, and production systems. Figure 1-3 and Figure 1-4 are taken from an actual SAP two-site landscape design and demonstrates the evolving complexity and proliferation of systems and servers.

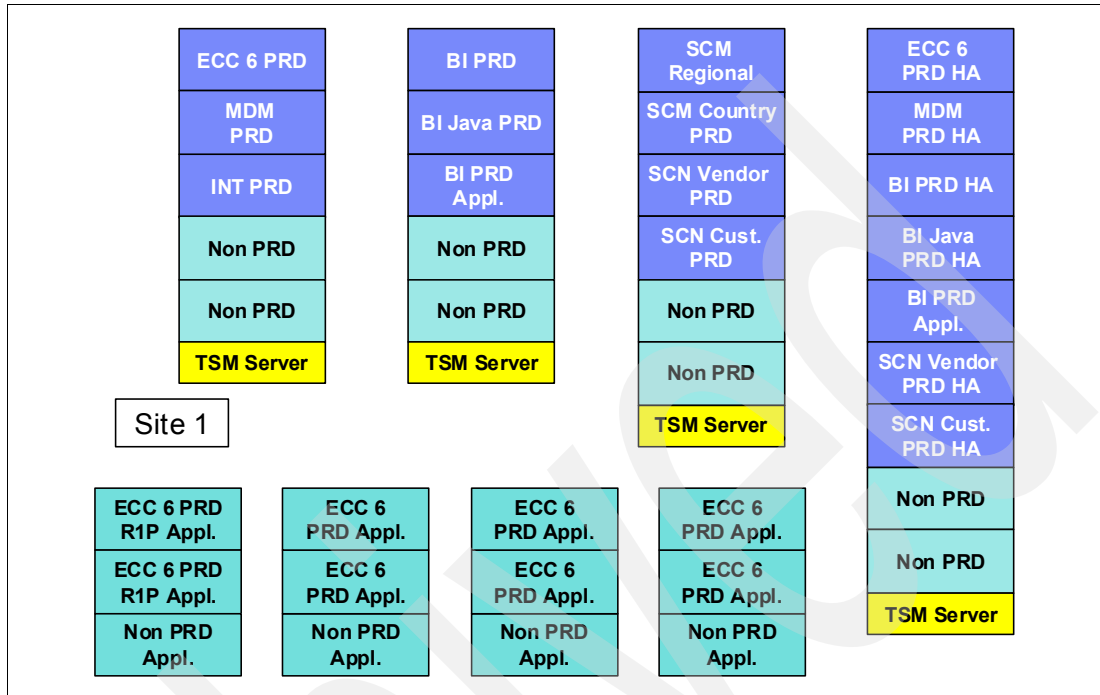


Figure 1-3 Site 1, multiple physical systems running multiple applications

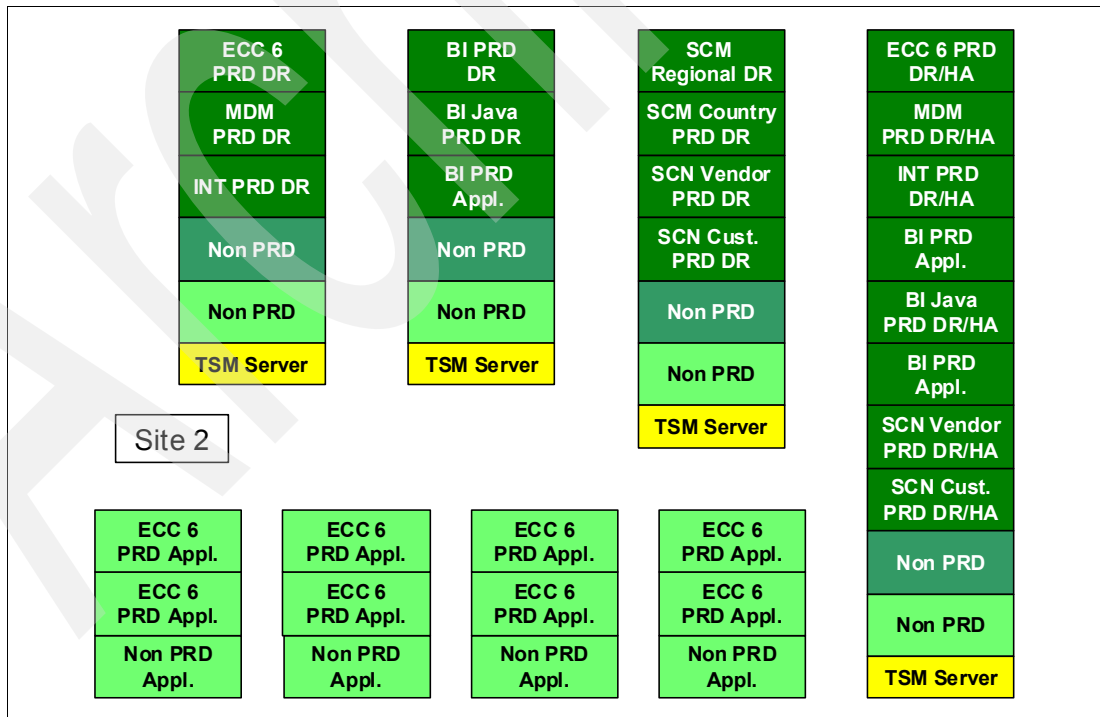


Figure 1-4 Site 2, multiple physical systems running multiple applications

Over the system life cycle and the production day, the utilization of these systems will vary considerably. Development and training are normally active during the day, quality assurance testing is a periodic activity, and production systems normally have different capacity requirements, depending on the time of day, day of month, and type of system. A number of studies show that the average processor utilization of a server in an SAP landscape is surprisingly low, in the order of 15-20%. With each system residing on its own dedicated server (or group of servers in 3-tier), idle resources remain dedicated to idle systems.

### **1.1.2 Motivation for server virtualization**

There are a number of motivations for moving to virtualized server resources, the primary focus for all of them is on reduction of complexity and the total cost of ownership of the IT landscape, which you can realize through consolidation to achieve a more efficient utilization of hardware resources. SAP systems are normally sized for the capacity that is required to meet the critical peak workload, which often represents only a small part of the system's productive day, month, or year. By consolidating a large number of dedicated servers, each running at an average low utilization with only periodic peaks, into a shared infrastructure, less capacity is required in total. This consolidation normally reduces the system management and maintenance overhead too by reducing the number of servers. A recent aspect to this calculation, which plays an ever increasing role, is also the fact that server consolidation can also reduce the number of servers that idly consume power and space.

### **1.1.3 Beyond the server**

Each SAP system has an associated storage requirement because each has at least one database. The proliferation in the direction of the storage infrastructure reflects the same type of complexity that you see in the landscape itself. In this area, the move from dedicated storage is replaced by shared storage servers, and a networked storage infrastructure becomes the norm. Nevertheless, most servers have their own dedicated access paths and statically dedicated storage capacity that generates maintenance overhead per system. The use of virtual I/O servers (which allow sharing of the I/O access paths) and virtual storage implementations (which provide a layer of abstraction between the logical storage view and the actual storage hardware) begin to address this complexity. These solutions help to reduce scheduled downtime and provide a high degree of flexibility in storage maintenance. We do not cover storage virtualization in this book.

## **1.2 Examples of virtualization advantages**

We use the four examples in this section to demonstrate a variety of situations in which virtualization is used to a proven advantage.

### **1.2.1 Landscape consolidation**

There are multiple examples where customers consolidated dozens or even hundreds of systems, running on individual physical servers, into a few large servers. These few servers can then carry all of the consolidated workload, which was previously distributed over the many small servers, each with reserve capacity to cover an individual workload peak. This consolidation reduced the number of servers and the need for idle capacity and cost for power and cooling. Figure 1-5 on page 5 shows an example of the individual load profiles for various systems running on dedicated servers and the results of these loads when combined using processor sharing into a single hardware server.

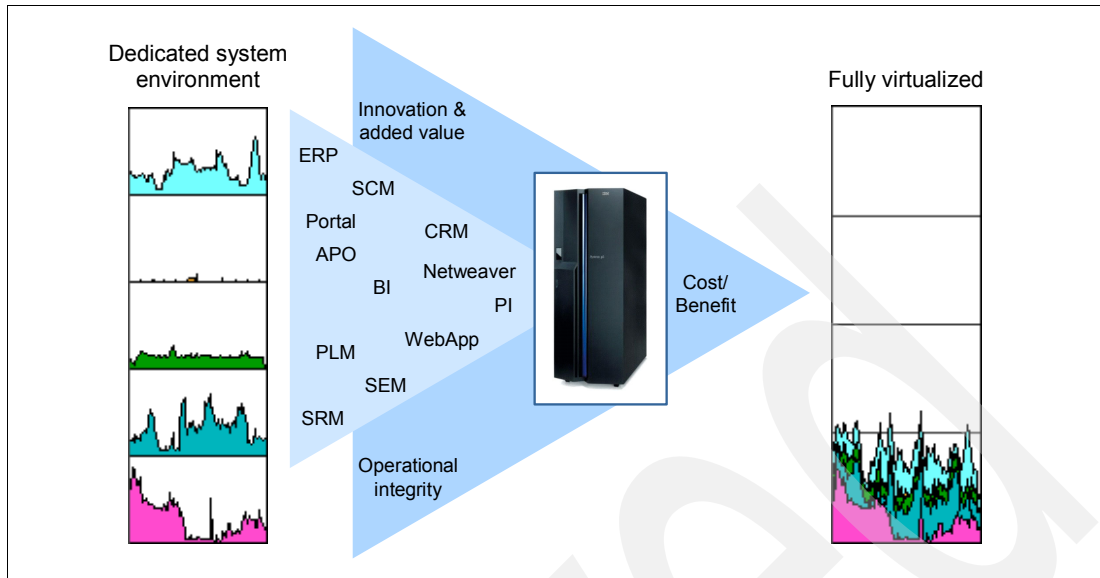


Figure 1-5 Possible consolidation scheme of an SAP system landscape

## 1.2.2 Load shifting

You can use virtualization to cope with shifting load requirements during the processing life cycle. You find this type of behavior in an integrated SAP landscape, where business processes span a number of SAP systems, shifting the processing load from one system to another. By using virtualization, the resources can follow the load shift and meet the peak load in each of the SAP systems. Compared to dedicated servers, where each server is individually sized to cover the peak load of one SAP system, the same implementation in a virtualized landscape needs to cover only the highest peak at any point-in-time, which can provide a big benefit.

Figure 1-6 shows an example of this type of consolidation in the business processes for service parts management (SPM) for automotive. Here the business processes for the order fulfilment span three SAP systems: CRM, SCM, and ECC. Each unit-of-work executes a process sequence that drives the workload in each of the three systems. The result is simultaneous workload on all three servers.

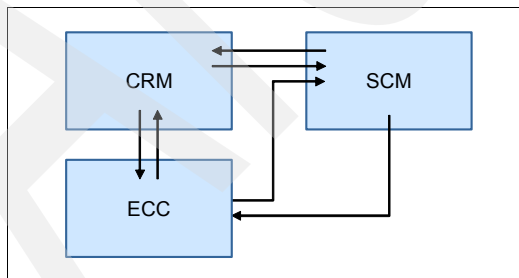


Figure 1-6 Example of the components of an SAP service parts management solution in the automotive industry

Figure 1-7 on page 6 shows measurements of the peak and average processor requirements during the order fulfilment run for a high-volume system test. Individually SCM needed a maximum of 31 processors. ECC needed 16 processors and CRM 28 processors. In this case, to cover the peak requirements of all components in a dedicated server landscape, 76

processors are required. Using the process in Figure 1-7, processor virtualization covered all peaks for each system in this integrated environment with 64 processors.

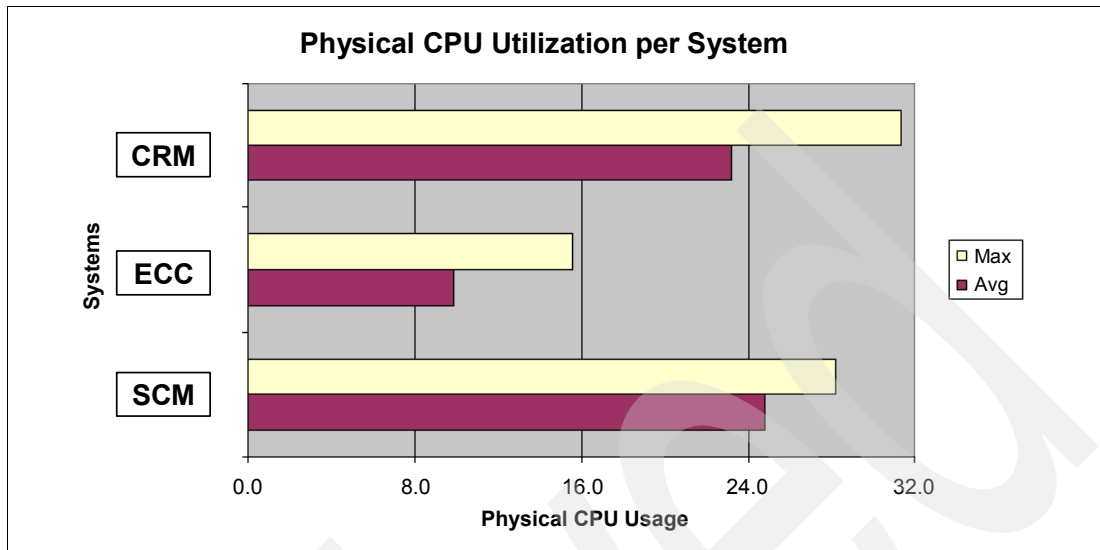


Figure 1-7 Distribution of processor utilization in the service parts management scenario

Figure 1-8 shows the capacity utilization for the same order fulfilment run that Figure 1-7 shows, as it looks when recorded in two minute intervals. The maximum or peak utilization (the y-axis shows processor utilization in %) for each of the systems is averaged over two minutes in this graph. The distribution of the load over the three systems appears to be a homogenous pattern.

The resource distribution in the shared processor pool is implemented in 10 ms time slices. The result of this fine granularity of the resource sharing is that even heavily competing peak loads gain considerable benefit.

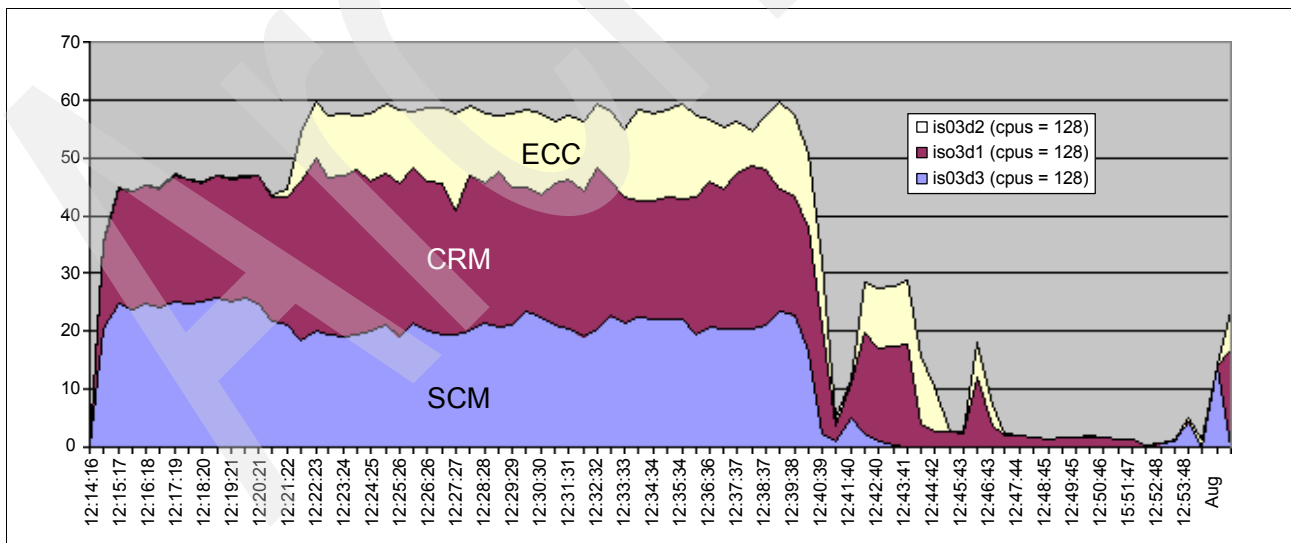


Figure 1-8 Monitoring snapshot in two minute intervals

If the finest granularity of the monitoring tool is used to observe the load behavior of the SAP systems, it becomes clearer how this is achieved. Figure 1-9 on page 7 is another example of



the workload running on the SCM and ECC systems in the same SPM scenario of Figure 1-8 on page 6, but in a separate test.

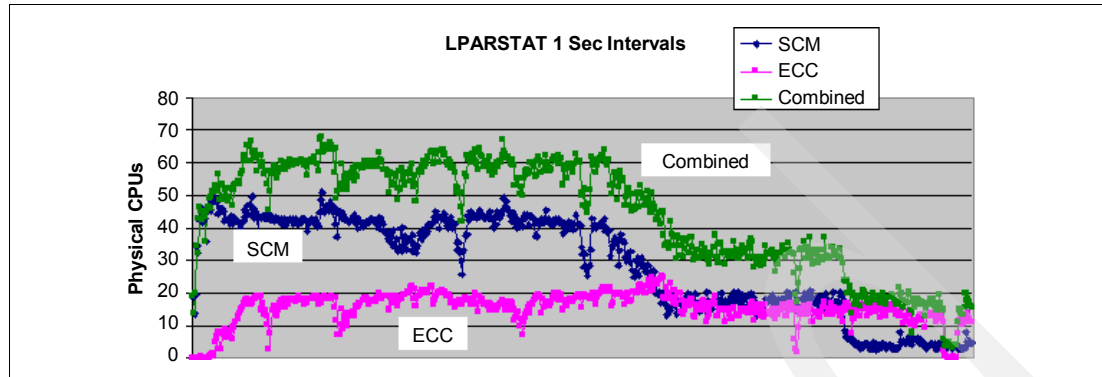


Figure 1-9 Processor utilization from lparstat in one second intervals

Figure 1-9 is taken from the high-load test, which required the CRM system to be moved to a second p595. In this graph, we track only the ECC and SCM systems that share this physical machine.

The measurements are taken in one second intervals (every mark on the graphs is a one second average). The “combined” line represents the sum of the peaks at the given interval for both SCM and ECC. In this graph, we see how the volatile load peaks mesh together. The combined load, evaluated in the short one second interval, reaches 68 processors. Because there are only 64 processors in the system, this shows that the peaks occur at an even more granular level than one second at different times, hence the two workloads balance each other out.

If we drilled down and viewed the systems in 10 ms intervals, we might see the further intensity of this load fluctuation.

### 1.2.3 Processing chain

There is also a benefit in virtualization for a multi-step processing chain that has different requirements at each stage of processing. In this case, the highest peak must be covered, but the resources not required by the less intensive steps are shared with the other workload.

Figure 1-10 is taken from a proof-of-concept for road billing implementation. In this scenario, there are a total of 15 processing steps that are executed in a sequential chain, nine of which are documented in this proof of concept.

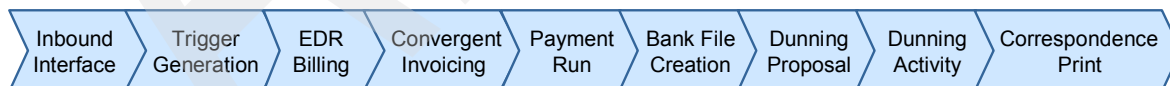


Figure 1-10 End-to-end business process chain for “Service to cash”

Figure 1-11 on page 8 shows the processor requirements of each of the sequential steps in Figure 1-10, broken down in requirements for the database and for the application servers. During the high-volume processing run in this utilization chart, several of the critical steps require significant processor power and a number are comparatively light. To complete the work within the required batch processing window, the machine capacity must cover the peaks of the high load steps, which results in a good deal of idle capacity, even during the critical batch processing window. This is capacity potentially available for other workload.

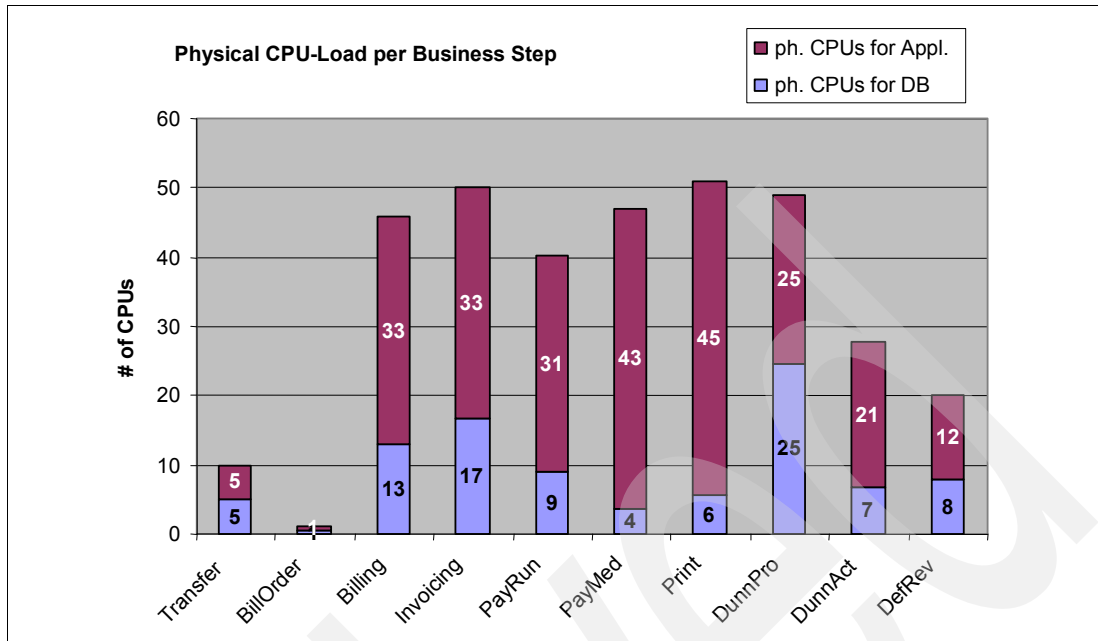


Figure 1-11 Processor utilization for the different business processes

Additionally, let us consider the shift in requirements over these steps in the distribution of the resources between the database and application servers. This shift is as great as 7:1 (application server processor to DB server processor ratio) in Correspondence Printing (Print) versus 1:1 in Dunning Proposal Activities (DummPro). The ECC system that runs this load is implemented as a 3-tier configuration, which occurred in consideration of the high-availability fail-over scenario for the database and central instance. In this case, the database and the production application servers are in two different LPARs. In a dedicated 3-tier architecture of this design, for the sizing consider the peak requirements of each of these components. In this case, the requirement is for 70 processors (rather than the 51 used or nearly 40% more capacity) because the application peak is at 45 processors and the DB peak is 25 processors, albeit at different stages of processing.

## 1.2.4 Concurrent load prioritization

Many SAP systems are not dealing with a single load type, but concurrent loads of different profiles and priorities. The mix often comprises online transactional load, batch processing, and administration activities, such as inter-system data transfers or online database backups. Virtualization can also be used here to separate and prioritize the load types according to the business requirements to take full advantage of the available processing capacity.

As an example, Figure 1-12 on page 9 is taken from a high-end SAP Business Intelligence proof-of-concept. This system represents a 24x7 “follow the sun” implementation with four primary workload considerations: online transaction processing, batch data loading, aggregation of loaded data, and administration. In this implementation, the load types were separated into dedicated application server LPARs, which were prioritized using a combination of guaranteed and shared resources. Using a simple-to-implement scheme, the batch processing can utilize all “left over” processing cycles without measurable impact to the online or database activities. Using this approach, batch loading can continue at full throttle concurrent with online user activities.

Additionally, resources that are normally reserved for administration activities, such as the “data movers” that perform the backup of a system flash-copy to tape each eight hours, are

available for production activities outside of the (relatively small) backup window. They are guaranteed to the data movers but available to other shared processor LPARs (SPLPARs, for details see 2.9, “Micropartitioning and Shared Processor LPARs” on page 17) when not in use.

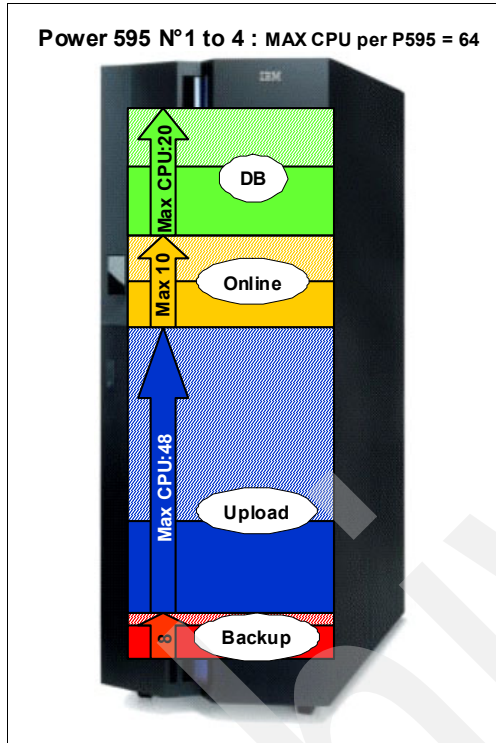


Figure 1-12 Power Systems 595 with a schema of the various load types

These are a few examples of where you can use virtualization to achieve the best utilization of resources, reduce unnecessary idle capacity, reduce the landscape hardware distribution, and reduce the total cost-of-ownership. In the next chapters, we describe the technology in detail and position the virtualization technology into the SAP landscape.

Archived

## PowerVM virtualization technologies

Over the past few years, IBM has introduced a number of virtualization technologies, capabilities and offerings. Many of the virtualization features were originally available through an offering known as Advanced POWER Virtualization (APV). As of early 2008, these and many subsequent technologies and offerings have been collectively grouped under the new brand of PowerVM.

The PowerVM technologies are packaged into three PowerVM editions:

- ▶ PowerVM Express Edition
- ▶ PowerVM Standard Edition
- ▶ PowerVM Enterprise Edition

These packages give you the flexibility to choose a cost-effective solution based on your virtualization requirements.

PowerVM is additionally extended by virtualization features and capabilities of the processor and operating system.

Individual technologies are enabled either by hardware, software, firmware, or some combination thereof. Table 2-1 presents a list of key technology offerings with their requirements.

Table 2-1 PowerVM technologies by PowerVM Edition and technology enabler

PowerVM Technology	Express Edition	Standard Edition	Enterprise Edition	Hardware requirement <sup>a</sup>	Available with AIX <sup>b</sup>	Available with IBM POWER Linux <sup>b</sup>	Available with IBM i <sup>b</sup>
Hypervisor	✓	✓	✓	POWER4™	✓	✓	✓
Hardware Management Console		✓	✓	POWER4	✓	✓	✓
Integrated Virtualization Manager	✓	✓	✓	POWER4	✓	✓	✓
Systems Director Management Console	✓	✓	✓	POWER6®	✓	✓	✓
Systems Director VMControl™	✓	✓	✓	POWER5™	✓ AIX 5.3	✓	✓ c
Dedicated Logical Partitions (LPARs)	✓	✓	✓	POWER4	✓	✓	✓
Live Partition Mobility			✓	POWER6	✓ AIX 5.3	✓ SLES10 SP2, RHEL5 QU2	d
Dynamic LPAR	✓	✓	✓	POWER4	✓	✓	✓
Micro-Partitioning™ and Shared Processor LPARs	✓	✓	✓	POWER5	✓ AIX 5.3	✓	✓
Shared Dedicated Capacity	✓	✓	✓	POWER6	✓	✓	✓
Multiple shared Processor Pools		✓	✓	POWER6	✓	✓	✓
Virtual I/O Server (VIOS)	✓	✓	✓	POWER5	✓	✓	✓ IBM i 6.1
Partition Suspend / Resume		✓	✓	POWER7™	✓ AIX 6.1		✓ IBM i 7.1 TR2
N Port ID Virtualization	✓	✓	✓	POWER6	✓ AIX 5.3	✓ SLES 11 RHEL 6	✓ IBM i 6.1.1
Virtual Tape	✓	✓	✓	POWER6	✓ AIX 5.3	✓ SLES 11 RHEL 6	✓
Virtual SCSI	✓	✓	✓	POWER5	✓ AIX 5.3	✓	✓

PowerVM Technology	Express Edition	Standard Edition	Enterprise Edition	Hardware requirement <sup>a</sup>	Available with AIX <sup>b</sup>	Available with IBM POWER Linux <sup>b</sup>	Available with IBM i <sup>b</sup>
Virtual Ethernet	✓	✓	✓	POWER5	✓ AIX 5.3	✓	✓
Shared Ethernet Adapter		✓	✓	POWER5	✓ AIX 5.3	✓	✓
Integrated Virtual Ethernet	✓	✓	✓	POWER6	✓	✓	✓
Active Memory™ Sharing (AMS)			✓	POWER6	✓ AIX 6.1	✓ SLES 11 RHEL6	✓ IBM i 6.1.1 + PTFs
Active Memory Expansion (AME)	n/a <sup>e</sup>	n/a <sup>e</sup>	n/a <sup>e</sup>	POWER7	✓ AIX 6.1		
Workload Partitions	✓	✓	✓	POWER4	✓ AIX 6.1		
Workload Partition Manager	✓	✓	✓	POWER4	✓ AIX 6.1		
Live Application Mobility	✓	✓	✓	POWER4	✓ AIX 6.1		
Simultaneous Multithreading	✓	✓	✓	POWER5	✓	✓	✓
IBM i Subsystems	✓	✓	✓	any			✓

a. Minimum hardware technology

b. Minimum software release level

c. IBM i only supports features of the SystemsDirector VMControl Express Edition

d. Statement of Direction: Live Partition Mobility is planned to be made available on IBM i with a Technology Refresh for IBM i 7.1. This support will require POWER7.

e. Not related to PowerVM editions

Table 2-1 on page 12 shows the minimum requirements with respect to the hardware. Individual models might not support all features. Check the following website for details:

<http://www.ibm.com/systems/power/software/virtualization/editions/index.html>

## 2.1 Hypervisor

The POWER Hypervisor™ is a foundation technology for PowerVM virtualization. It exists as a firmware layer between the hosted operating systems and the hardware and provides functions that enable many of the PowerVM technologies. These are, for example, dedicated logical partitions (LPARs), micro-partitions, shared processor pools, dynamic LPAR reconfiguration, virtual I/O, and virtual LAN. For example, as seen in Figure 2-1 on page 14, for managing partitions, the Hypervisor dispatches partition workloads amongst the processors, ensures partition isolation, and supports the dynamic resource movement.

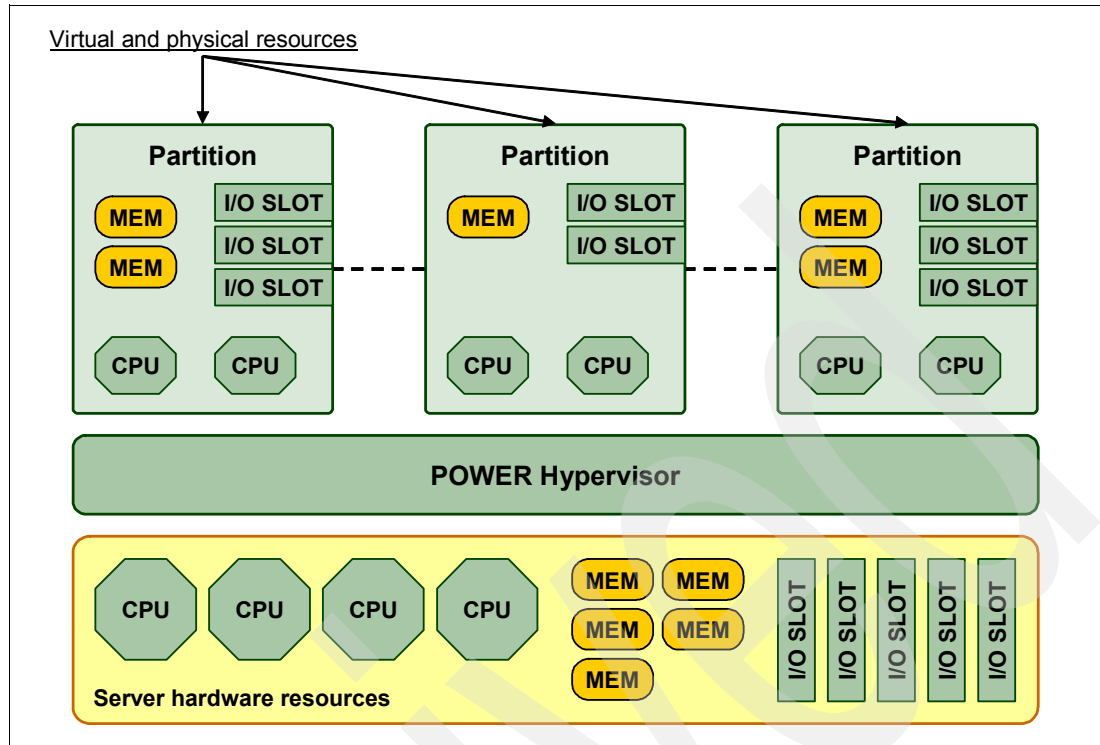


Figure 2-1 Power Hypervisor

The POWER Hypervisor is always installed and activated, regardless of the system configuration. Although the POWER Hypervisor has no specific or dedicated processor resources assigned to it, it does consume a small overhead in terms of memory and processor capacity from both system and LPAR resources.

The managing interface to the Hypervisor is either the Hardware Management Console (HMC), the Integrated Virtualization Manager (IVM), the Systems Director Management Console (SDMC), or Systems Director VMControl.

## 2.2 Hardware Management Console

The Hardware Management Console (HMC) is a dedicated Linux-based appliance that you use to configure and manage IBM Power System servers. The HMC provides access to logical partitioning functions, service functions, and various system management functions through both a browser-based interface and a command line interface (CLI). Because it is a separate stand-alone system, the HMC does not use any managed system resources and you can maintain it without affecting system activity.

## 2.3 Integrated Virtualization Manager

For smaller or segmented and distributed environments, not all functions of an HMC are required, and the deployment of an additional management server might not be suitable. IBM developed an additional hardware management solution called the Integrated Virtualization Manager (IVM) that provides a convenient browser-based interface and can perform a subset of the HMC functions. It was integrated into the Virtual I/O Server product that runs in a



separate partition of the managed server itself, which avoids the need for a dedicated HMC server. Because IVM itself is provided as a no cost option, it lowers the cost of entry into PowerVM virtualization. However, the IVM can only manage a single Power System server. If IVM is the chosen management method, then a VIOS partition is defined and all resources belong to the VIOS. That means that no partition that is created can have dedicated adapters; instead, they are all shared.

## 2.4 Systems Director Management Console

As of April 2011, IBM has released the Systems Director Management Console (SDMC) as the next generation HMC. SDMC extends the support scope of the HMC to range from POWER blade servers to the high-end Power servers, thereby unifying and consolidating administration. It also leverages the same consistent Systems Director user interface utilized for other IBM hardware ranging from mainframes to System x® servers. To ease the transition, SDMC can operate simultaneously with existing HMC or IVM consoles. It is currently offered on the same x86-based hardware as the HMC, as well as being available as a Red Hat Enterprise Virtualization KVM or VMware-based virtual appliance.

## 2.5 Systems Director VMControl

Another system management offering available for Power System servers is the Systems Director VMControl. VMControl is a cross-platform virtualization management solution providing an enterprise-wide management platform for servers, storage, networks, and software. It is offered as a virtualization plug-in for IBM Systems Director.

With the same dashboard interface, VMControl not only enables you to create, edit, manage and relocate LPARs, but you can also capture ready-to-run virtual LPAR images and store them in a shared repository. These images can be quickly deployed to meet business needs. You can also create and manage server and storage system pools to consolidate resources and increase utilization.

The available features and capabilities are packaged into three editions as depicted in Table 2-2: Express, Standard, and Enterprise.

Table 2-2 Express, Standard and Enterprise edition of VMControl

Feature	Express	Standard	Enterprise
Create and manage virtual machines	✓	✓	✓
Virtual machine relocation	✓	✓	✓
Import, edit, create and delete virtual images		✓	✓
Deploy virtual images		✓	✓
Maintain virtual images in a repository		✓	✓
Manage virtual workloads in system pools			✓

## 2.6 Dedicated LPARs

Logical partitioning was introduced to the Power Systems environment on POWER4-based servers. It provided the ability to make a server run as though it were two or more independent servers. When a physical system is logically partitioned, the resources on the server are divided into subsets called logical partitions (LPARs). Processors, memory, and input/output devices can be individually assigned to logical partitions. Dedicated LPARs hold these resources for exclusive use. You can separately install and operate each dedicated LPAR because LPARs run as independent logical servers with the resources allocated to them.

## 2.7 Live Partition Mobility

Live Partition Mobility (LPM) provides the ability to move a running AIX or Linux partition from one physical POWER6 technology-based system or newer server to another compatible server without application downtime. This feature allows applications to continue running during activities that previously required a scheduled downtime, for example, for hardware and firmware maintenance and upgrades, for workload rebalancing, or for server consolidation. For more details, see Chapter 6, “IBM PowerVM Live Partition Mobility” on page 63.

## 2.8 Dynamic LPAR

Dynamic logical partitioning (DLPAR) gives you the ability to manually move resources (such as processors, I/O, and memory) to, from, and between running logical partitions without shutting down or restarting the logical partitions.

When you apply this dynamic resource allocation, known as dynamic logical partitioning or dynamic LPAR, you can redefine all available system resources to reach optimum capacity for each partition, which allows you to share devices that logical partitions use occasionally. The following examples describe situations in which you might want to employ dynamic LPAR:

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

## 2.9 Micropartitioning and Shared Processor LPARs

Shared Processor LPARs (SPLPARs) are logical partitions that share a common pool of processors, which is called the shared-processor pool, as illustrated in Figure 2-2 on page 18. Micropartitioning technology allows these partitions to be sized using 1/100th of processor increments whereas SPLPARs can start as small as 1/10th of a processor. This level of processor granularity provides excellent flexibility (in comparison to dedicated processor LPARs) when allocating processor resources to partitions. Within the shared-processor pool, unused processor cycles can be automatically distributed to busy partitions on an as-needed basis, which allows you to *right-size* partitions so that more efficient utilization rates can be achieved. Implementing the shared-processor pool using micropartitioning technology allows you to create more partitions on a server, which reduces costs.

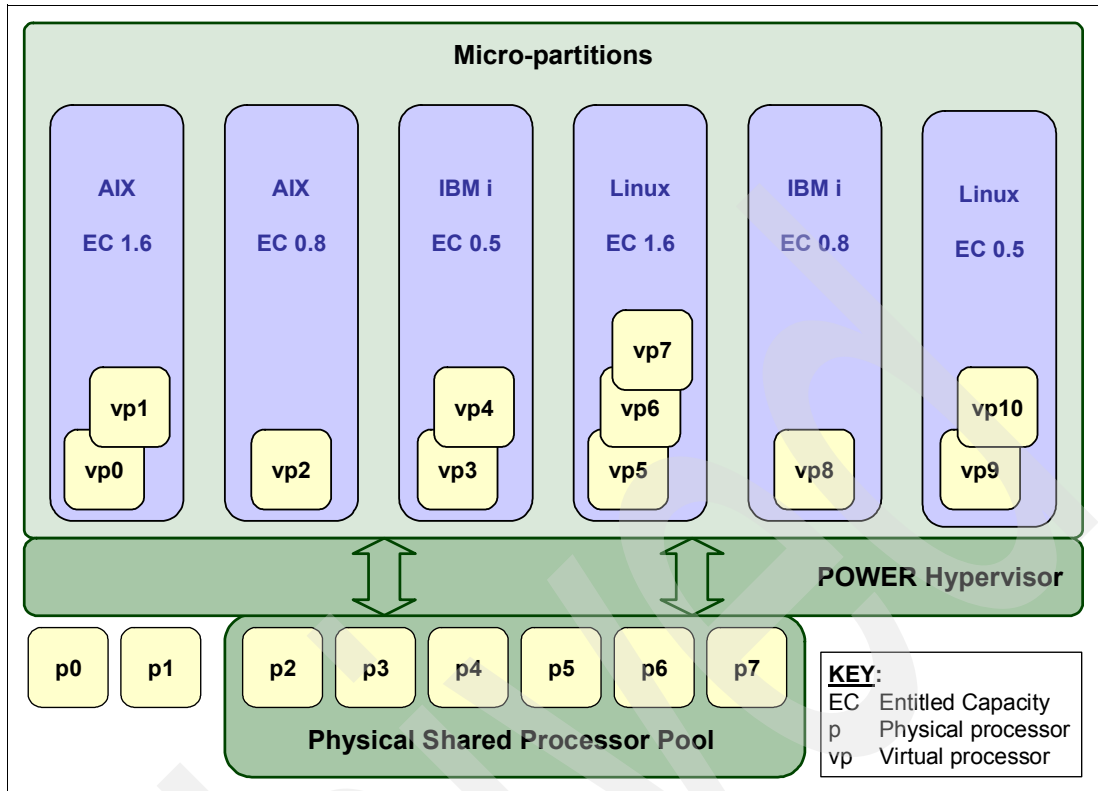


Figure 2-2 Micro-Partitioning and Shared Processor LPARs

## 2.10 Shared Dedicated Capacity

A new feature in POWER6 technology-based systems, Shared Dedicated Capacity, allows partitions that are running with dedicated processors to *donate* unused processor cycles to the shared-processor pool. When enabled in a partition, the size of the shared processor pool is increased by the number of physical processors that are normally dedicated to that partition, which increases the simultaneous processing capacity of the associated SPLPARs. Due to licensing concerns, however, the number of processors an individual SPLPAR can acquire is never more than the initial processor pool size. This feature provides a further opportunity to increase the workload capacity of uncapped micro-partitions.

## 2.11 Multiple Shared-Processor Pools

Multiple Shared-Processor Pools (MSPP) is a feature that is available starting with POWER6 technology-based systems. It allows you to create additional shared-processor pools. The MSPP pools are subsets of and contained within the global shared-processor pool. You can then assign SPLPARs processing capacity from either the global shared processor pool or a newly created MSPP pool. The main motivation here is to limit the processor capacity for SPLPARs, thereby reducing software license fees that are based on processor capacity.

## 2.12 Virtual I/O Server

The Virtual I/O Server (VIOS), as illustrated in Figure 2-3, is an AIX-based appliance partition that provides access to its physical storage and network resources to virtual I/O client partitions. Client partitions can be either AIX, IBM i or Linux based. By eliminating the need for dedicated resources, such as network adapters, disk adapters and disk drives, for each partition, the VIOS facilitates both on demand computing and server consolidation. The VIOS is a foundation technology required by many other PowerVM technologies.

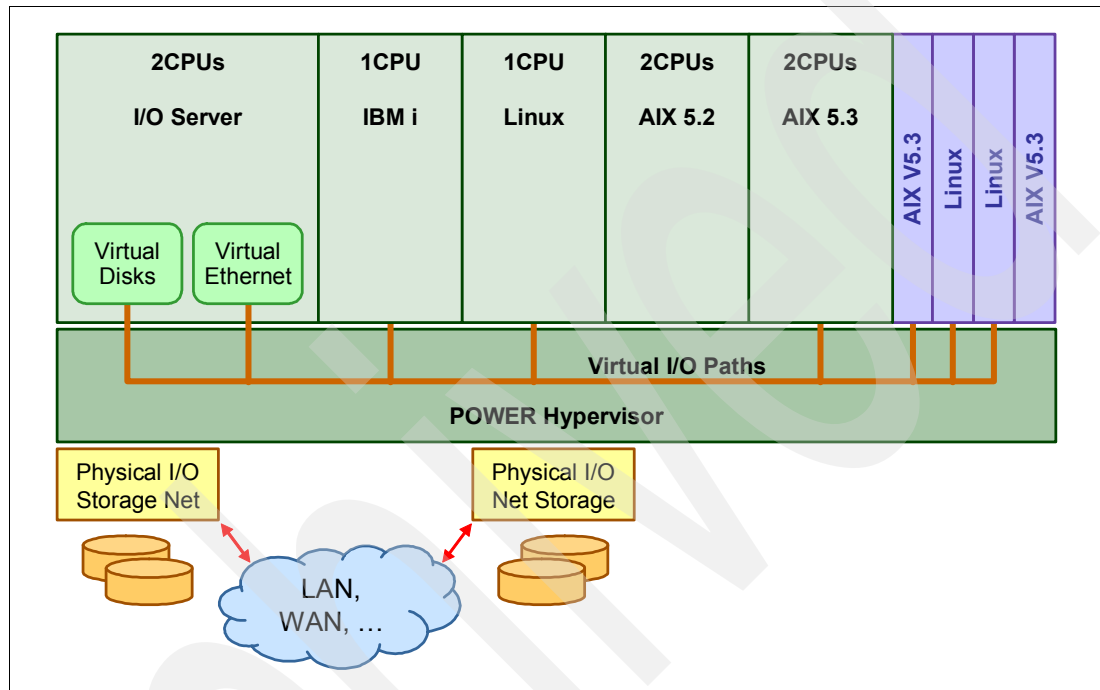


Figure 2-3 Virtual I/O Server

## 2.13 Partition Suspend and Resume

Partition Suspend and Resume provides the ability for partitions to be put into a standby or hibernated state and later to be restored and resumed back to their active state. During suspension, partition state information is stored on a VIOS-managed paging device.

A benefit of this is that partitions that are temporarily not required online can be suspended, freeing up the resources for other partitions.

Additionally, since some applications can have long shutdown and startup times, suspending and resuming a partition may be quicker than shutting down the LPAR and all its running applications and later restarting. In such situations, planned downtime for certain maintenance activities may also be reduced by avoiding a long shutdown and startup process.

The feature can also work in conjunction with LPM, to resume the partition on different systems.

## 2.14 N Port ID Virtualization

N Port ID Virtualization (NPIV) is an industry standard technology for virtualizing a physical Fibre Channel port. It provides the capability to assign a physical Fibre Channel adapter to multiple unique N Port IDs. Together with the Virtual I/O Server (VIOS) adapter sharing, this enables direct access to a Fiber Channel adapter from multiple client partitions. NPIV offers many benefits:

- ▶ Ease-of-use allowing storage administrators to use existing tools for storage management (including SAN managers, copy services, and backup and restore)
- ▶ Simplified storage provisioning using standard zoning and LUN masking techniques
- ▶ Physical and virtual device compatibility
- ▶ Access to SAN devices including tape libraries
- ▶ Distributed solutions that depend on SCSI heuristics SCSI-2 (Reserve/Release and SCSI3 Persistent Reserve)

## 2.15 Virtual Tape

Virtual Tape virtualizes physical Serial Attached SCSI (SAS) tape devices. Together with the VIOS, this allows the sharing of these devices among multiple client partitions.

## 2.16 Virtual SCSI

A virtualized implementation of the SCSI protocol is provided through the Virtual I/O Server. In large environments, the cost of adapters, switches, cables, patch panels, and so on can be a significant amount.

Virtual SCSI reduces the costs of provisioning storage to servers by sharing storage attachment costs among multiple partitions.

Additionally, Virtual SCSI and Virtual I/O might provide attachment of previously unsupported storage solutions. If the Virtual I/O Server supports the attachment of a storage resource, any client partition can access this storage by using virtual SCSI adapters

Virtual SCSI is based on a client/server relationship, as Figure 2-4 on page 21 illustrates. The Virtual I/O Server owns the physical resources and acts as server or, in SCSI terms, target device. The client logical partitions access the virtual SCSI backing storage devices that the Virtual I/O Server provides.

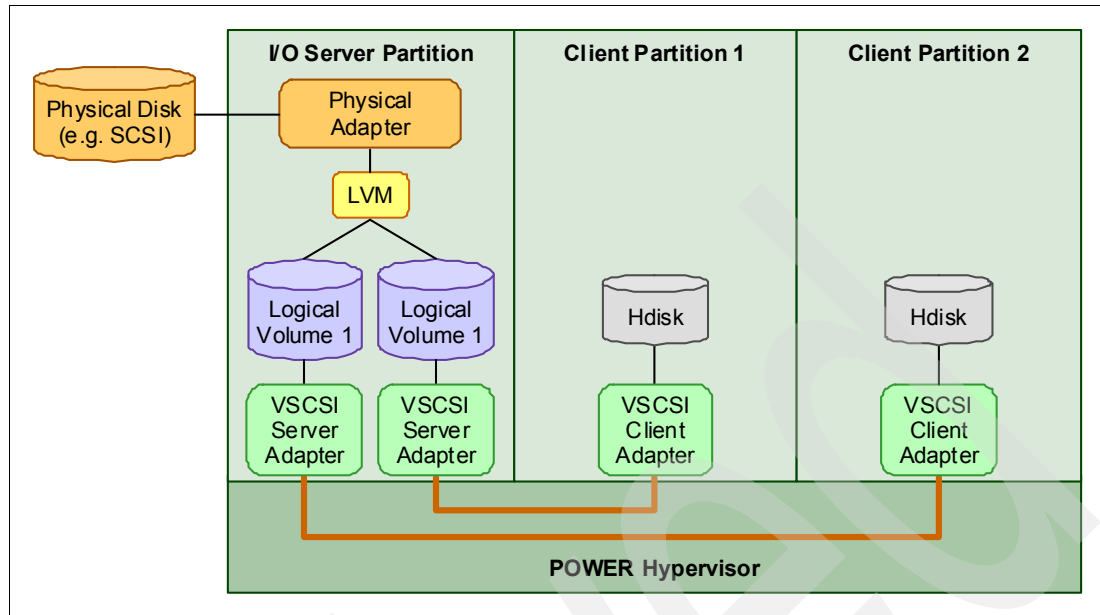


Figure 2-4 Virtual SCSI

## 2.17 Virtual Ethernet

Virtual Ethernet enables inter-partition communication without the need for physical network adapters assigned to each partition. In-memory connections between partitions are established and handled at the system level (Hypervisor and operating system interaction). Figure 2-5 on page 22 shows an example of two virtual LANs that are connected using virtual Ethernet adapters. These connections exhibit characteristics that are similar to physical high-bandwidth Ethernet connections and support standard industry protocols (such as IPv4, IPv6, ICMP, and ARP).

Virtual Ethernet requires at least a POWER5 technology-based system and the appropriate level of the AIX (5.3 onwards), IBM i or Linux operating systems.

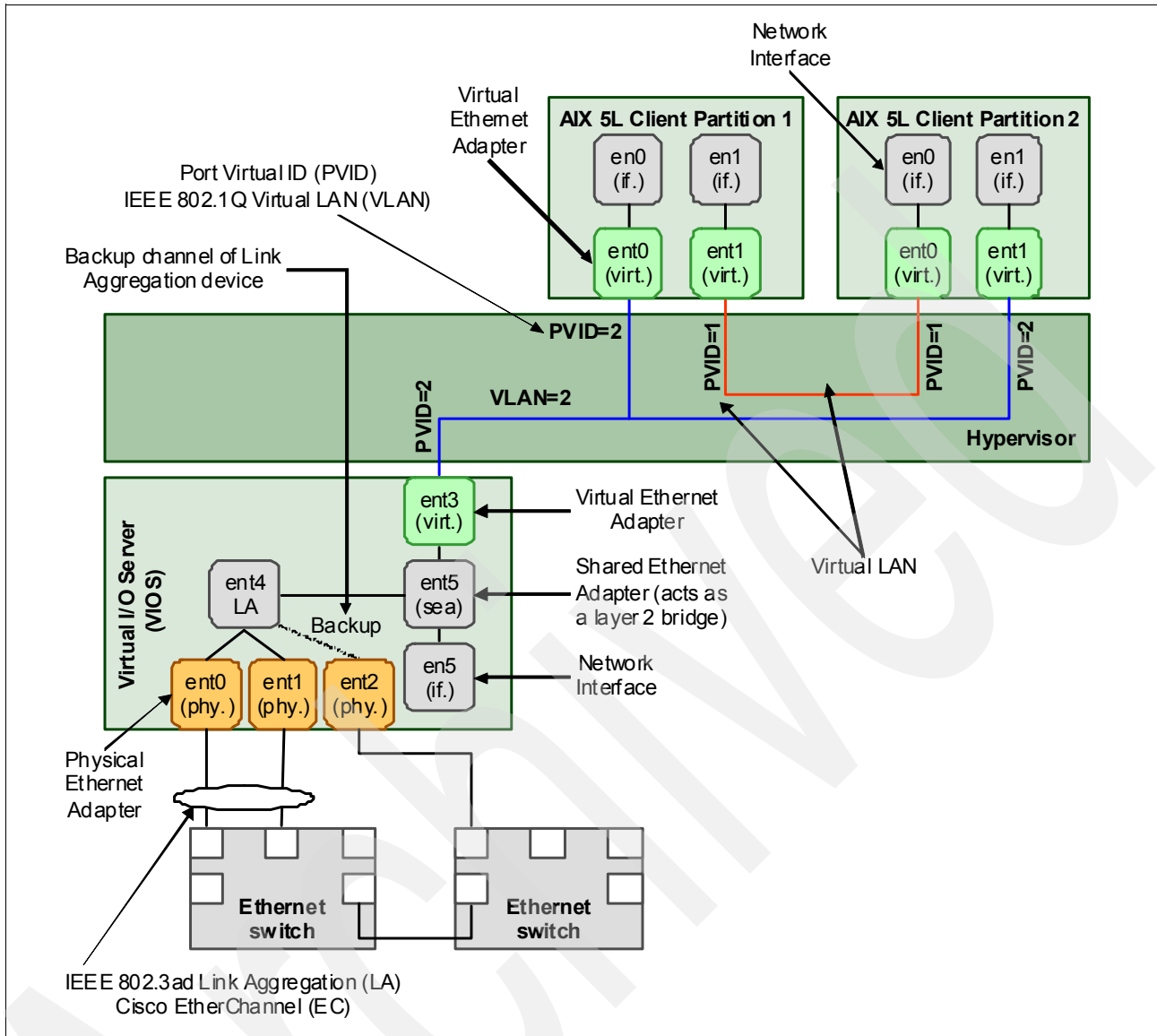


Figure 2-5 Virtual Ethernet

## 2.18 Shared Ethernet Adapter

Shared Ethernet Adapters enable multiple partitions to share a physical adapter for access to external networks, which allows partitions to communicate outside of the system without requiring the partition to dedicate a physical I/O slot with a physical network adapter. A Shared Ethernet Adapter (SEA) is created on the VIOS and acts as a bridge between a virtual Ethernet network and a physical network. For additional bandwidth or redundancy purposes, a SEA can utilize and aggregate multiple physical ports.



## 2.19 Integrated Virtual Ethernet

The Integrated Virtual Ethernet adapter (IVE) is a feature that is provided with POWER6 technology-based systems. It provides external network connectivity for LPARs using dedicated ports without the need of a Virtual I/O Server. IVE is comprised of the Host Ethernet Adapter (HEA) and the software support. The HEA is a physical Ethernet adapter that is connected directly to the GX+ bus of a POWER6 (or later) processor-based server to provide high throughput and low latency connectivity. Virtualized logical ports called Logical Host Ethernet Adapters (LHEAs) are configured and directly associated with LPARs, as Figure 2-6 shows.

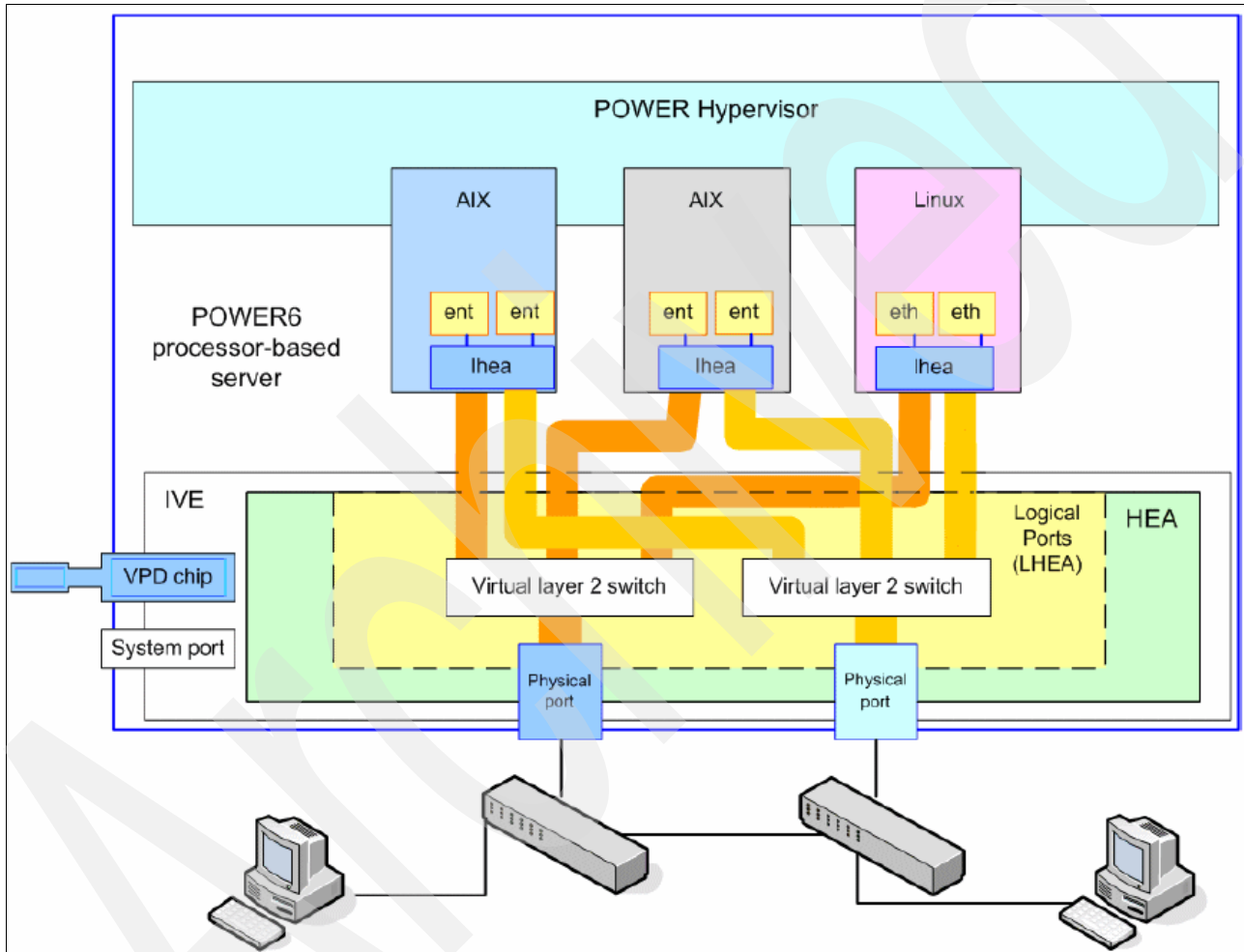


Figure 2-6 Integrated Virtual Ethernet adapter

IVE can provide most of the functions of both Virtual Ethernet and Shared Ethernet Adapters at improved performance and without the resource overhead of a Virtual I/O Server. However, some limitations do exist with IVE, such as the inability to perform LPM operations.

## 2.20 Active Memory Sharing

Active Memory Sharing (AMS) is a feature of PowerVM which allows for the sharing of system memory via a single physical memory pool amongst a set of LPARs (Figure 2-7 on page 24).

It is analogous to the sharing of processors from a processor pool by SPLPARs and the intent is to allow memory hungry partitions in a system to use portions of the physical memory not currently being used by other partitions.

When a partition is started, the configured memory defines the amount of logical memory assigned to a partition. The hypervisor then maps a range of physical memory to the partition's logical memory. In a dedicated memory partition this assignment remains fixed. In an AMS environment, on the other hand, the physical memory is part of a shared pool and portions are alternatively mapped to different AMS-managed LPARs' logical memory. Memory savings and optimization are achieved when the overall amount of memory resources are oversubscribed, that is, when the sum of the AMS-managed LPARs' logical memory is greater than the physical memory in the pool. To accommodate this, a VIOS-configured paging device is used to page out logical memory.

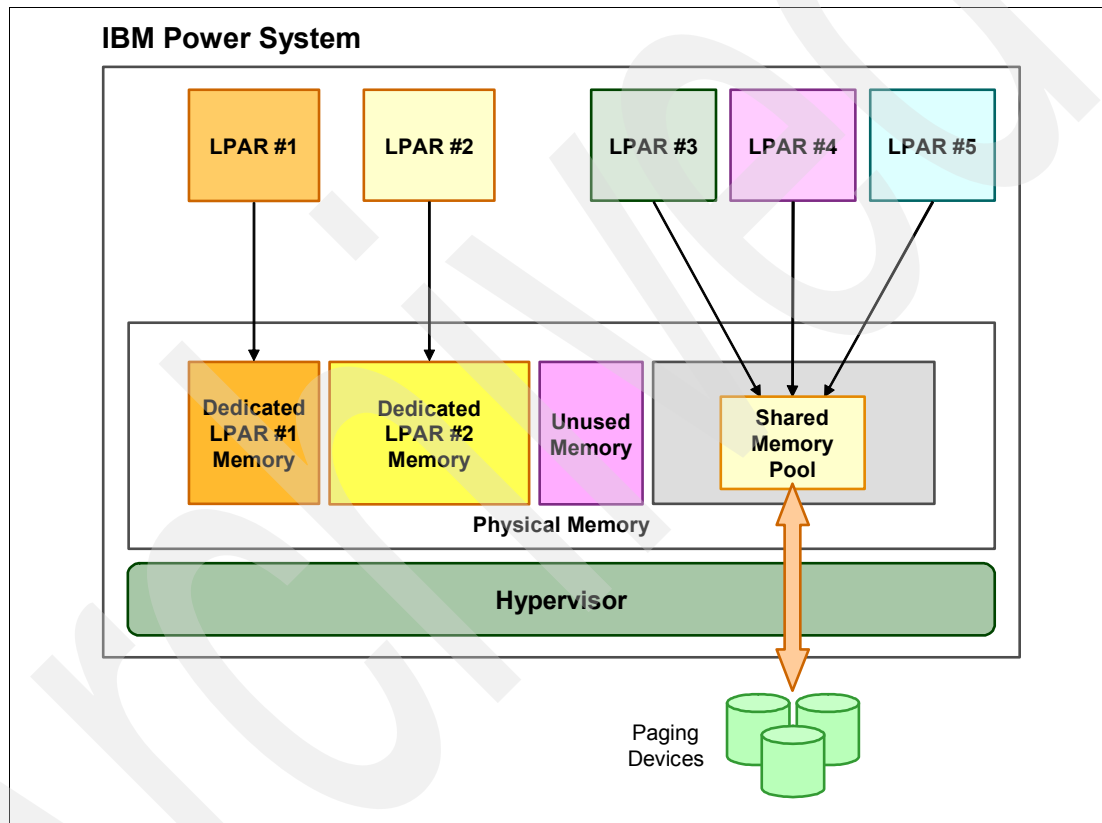


Figure 2-7 Shared and dedicated memory logical partitions

## 2.21 Active Memory Expansion

Beginning with the IBM POWER7 technology-based systems, LPARs with a minimum of AIX 6.1 TL4 SP3, can employ a new technology for expanding a system's effective memory capacity called Active Memory Expansion (AME). AME employs memory compression technology to transparently compress in-memory data, allowing more data to be placed into memory and thus expanding the memory capacity of the server. By utilizing Active Memory Expansion, clients can reduce the physical memory requirements or improve system utilization and increase a system's throughput. For details see 8.5, "Active Memory Expansion for SAP systems" on page 84.

## 2.22 Workload Partitions

Workload Partitions (WPAR) introduces virtualization of the operating system by providing isolated partitions of software services, applications, and administration within a single instance of the operating system. WPARs are a feature of the AIX operating system and are available as of AIX 6.1. Key benefits include rapid deployment and a reduction of the number of AIX images to maintain. For details see Chapter 7, “Workload partitions” on page 69.

## 2.23 Workload Partition Manager

The Workload Partition (WPAR) Manager provides a central systems management solution by providing a set of Web-based system management tools and tasks that simplify the management of a server and WPAR infrastructure, including Live Application Mobility. For details, see Chapter 7, “Workload partitions” on page 69.

## 2.24 Live Application Mobility

Live Application Mobility (LAM) is a WPAR feature that allows the WPAR to be relocated from one LPAR to another LPAR, running on the same or on separate physical servers. This feature helps to avoid planned downtime for supported applications that is caused by scheduled hardware maintenance. LAM might also help to improve performance by moving running instances to a more powerful server. Finally, it supports higher efficiency of servers in terms of a higher overall utilization and energy efficiency by concentrating applications on the best matching server capacity.

Live Application Mobility is not a replacement for high-availability software, such as PowerHA™ or similar products.

## 2.25 Simultaneous Multithreading

Simultaneous multi-threading (SMT) is a hardware feature that provides the ability for a single physical processor to simultaneously dispatch instructions from more than one thread context. For POWER5 and POWER6 technology-based processors, SMT enables two parallel threads (SMT2) while for POWER7 technology-based systems, up to four parallel threads per core are available (SMT4). For these processors, when SMT mode is activated, a single physical processor appears to the operating system as two logical processors in the case of SMT2 or four logical processors in the case of SMT4, independent of the partition type. For example, in SMT2 mode, a partition with one dedicated physical processor would operate with two logical processors. Similarly, in SMT4 mode, a shared processor partition with two virtual processors would appear as a logical 8-way. SMT is active by default.

The SMT performance effect is application dependent; however, most commercial applications see a significant performance increase. For SAP environments, it is recommended to have SMT turned on because the mix of many parallel online users, RFC, and batch tasks significantly benefits from this feature.

## 2.26 IBM i subsystems

In the IBM i operating system, subsystems are used to group processes and assign hardware resources in a flexible, yet controlled manner. In general, subsystems share all hardware resources of the logical partition and assign them to their processes based on the run priority of the processes and activity levels in the main storage pools. Subsystem properties are configured by subsystem description objects (object type: \*SBSD) and their related attributes and objects.

When running SAP on IBM i, each SAP instance is running in its own subsystem. By default, all SAP processes are running with the same run priority, each instance can use all available processor resources in the logical partition, and all SAP instances are sharing the main storage pool \*BASE. However, you can configure your SAP instances to run with different run priorities based on the work process types, to limit the total processor utilization of each instance, or to use different main storage pools for one or more instances. These configuration options are discussed in section 8.4, "Main storage pools, work process priorities, and workload capping on IBM i" on page 82.

## Best practice implementation example at a customer site

In the document “Banking IT Service Center running in a fully virtualized IBM PowerVM SAP environment”, a very efficient way of implementing a POWER-based SAP landscape is described in detail. You can get it here:

<http://www.ibm.com/software/success/cssdb.nsf/cs/STRD-8BKHRH?OpenDocument&Site=corp&ref=crdb>

The customer is using PowerVM technology since 2005 for the SAP landscape, which consists of roughly 90 SAP systems (all defined as 2-tier systems, DB, and CI). The four POWER systems are located in two data centers in a campus setup (distance roughly two kilometers). The overall average processor utilization of the physical systems was around 60% to 70%. The various applications were running each in their own SPLPAR and balanced each other out to achieve a high consolidation factor.

The lessons learned from this implementation can be summarized by the following few principles:

- ▶ For processor assignment: All LPARs are defined as shared and uncapped (an LPAR attribute that is configured on the HMC / Systems Director to allow the LPAR to grow up to the physical limit of the shared processor pool). The processor entitlement of the SPLPARs is defined as relatively small to allow headroom for the various SPLPARs that the POWER hypervisor manages.
- ▶ Virtual I/O servers (VIOS): Defined always in pairs, for load balancing and high availability reasons, running within the shared processor pool. Both virtual Ethernet and virtual disk I/O of all SPLPARs must use the VIO servers. For details, refer to Chapter 5, “Virtual I/O Server” on page 37.
- ▶ Memory: Because the memory is not assigned as flexibly as processor resources, there needs to be enough memory for each LPAR when it is expanded to its maximum. Expressed in popular scientific terminology, if the LPAR breathes in, it needs enough memory to process the increased workload properly. This is valid for systems, which do not use Active Memory Expansion (8.5, “Active Memory Expansion for SAP systems” on page 84) yet.

- ▶ **Combination of workloads:** The best consolidation effect can be achieved if everything is mixed and matched. That means that production, non-production, and different applications (for example, ERP, BI, CRM, SCM or Portal, and so on) are mixed on the same physical system, separated by individual LPARs for each SAP instance. The more and the smaller the SAP systems are, and the more varied their workload, the better the possible consolidation. The utilization peaks, as they are typically very short (“microspikes”), happen in the different systems at different times, which helps even more with the possible benefit of overall increased utilization.

Frequently, the question arises: If non-production and production systems are using the same shared processor pool, how do we ensure that the production systems do not get cannibalized by development, quality assurance, or sand-box systems? Our recommendation is to use the uncapped weight attribute of an LPAR. Give the VIO servers the highest weight. When the resources get tight, the VIOS still needs to serve all of the other LPARs on the system. Give the next weight level to the LPARs with the production systems, and then all the other LPARs. It does not make sense to use a too complex scheme of weight factors in one physical system. Three or four levels are sufficient for most cases.

You can get a good explanation of the weighting in the IBM Redpaper:

*IBM System p Advanced POWER Virtualization (PowerVM) Best Practices, REDP-4194*

In the remaining chapters, we provide more detailed recommendations and examples of hands-on activities for building a flexible and efficient landscape for an SAP IT environment.

## Hands-on management tasks

In this chapter, we provide some useful examples of HMC functionality, which ease the operations of virtualized SAP landscapes on PowerVM. If you are interested in further HMC capabilities, refer to:

*IBM PowerVM Virtualization Managing and Monitoring*, SG24-7590

<http://www.redbooks.ibm.com/abstracts/sg247590.html?Open>

### 4.1 Ease of moving from dedicated to shared

The change from dedicated to shared resources is simple. It requires that you make a slight change in the existing profile and restart the LPAR. Prior to doing this, of course, you must have a policy for sharing the physical server (CEC = Central Electronic Complex) resources that supports the processing requirements of the combined load, according to the business need. This policy determines the relative weights, the guaranteed entitlements, and the virtual processor (VCPU) distribution.

Set VCPU requirements to a value that reflects the expected peak workload requirements. The entitlement depends on how much capacity you want to guarantee, how many partitions are set up, and how many physical processors are available in the shared processor pool. The weight determines your relative processing priorities: online over batch, production over non-production, and so on. Using the dynamic tools that we discuss in this chapter, you can also fine-tune this policy during the operation to address a different load. You can get an understanding of the LPAR requirements for each SAP system from the SAP Computer Center Management System (CCMS), Insight reports (9.6, “IBM Insight for SAP” on page 157), and other monitoring tools.

To change an LPAR from a dedicated to a SPLPAR, you must create a new profile or change the existing profile. In our example, we alter the existing profile, save it, and restart the partition:

1. Login on the HMC, and access the Manage Profiles panel by selecting **Systems management** → **Custom Groups** → **All partitions**. Choose the partition to be changed.

2. To open the panel for managing profiles, select **Configuration** → **Manage Profiles**, as shown in Figure 4-1.
3. Select the profile to be changed, and choose **Actions** → **Edit**. Alternatively, create a copy, and then change the new profile, as shown in Figure 4-1.

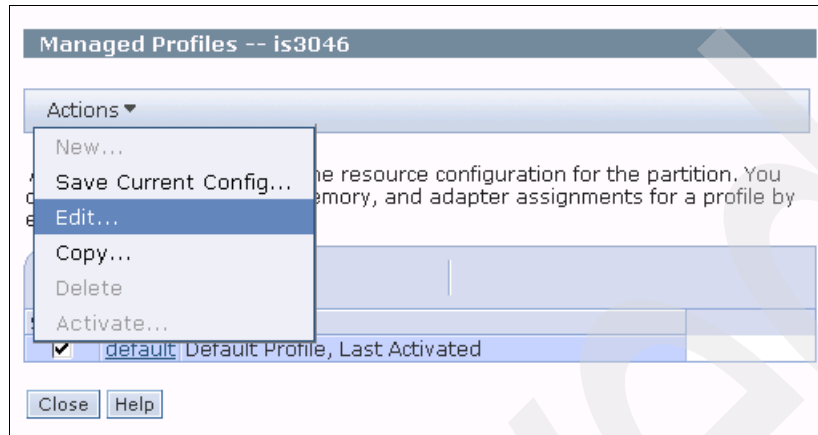


Figure 4-1 Select a profile to edit



- The profile opens, as shown in Figure 4-2. Select the **Processor** tab, select the **Shared** option, and change the resources accordingly to your requirements. Figure 4-3 on page 32 is an example of assigning resources.

The shown LPAR can later use up to one physical processor if enough resources in the shared pool are available with what corresponds to the setting in the dedicated setup before. In addition, we allow an increase to the amount of processor without a restart, for example when using scheduled DLPAR operations or adding new hardware (Maximum processing units and Maximum virtual processors). To later benefit from the implicit capping, which we introduce in 4.3, “Assigning implicit capping to LPARs” on page 33, we create it as uncapped, which is the recommended partition mode and type used for SAP environments with PowerVM.

- Shut down the LPAR, and start it up again by activating the new or altered profile.

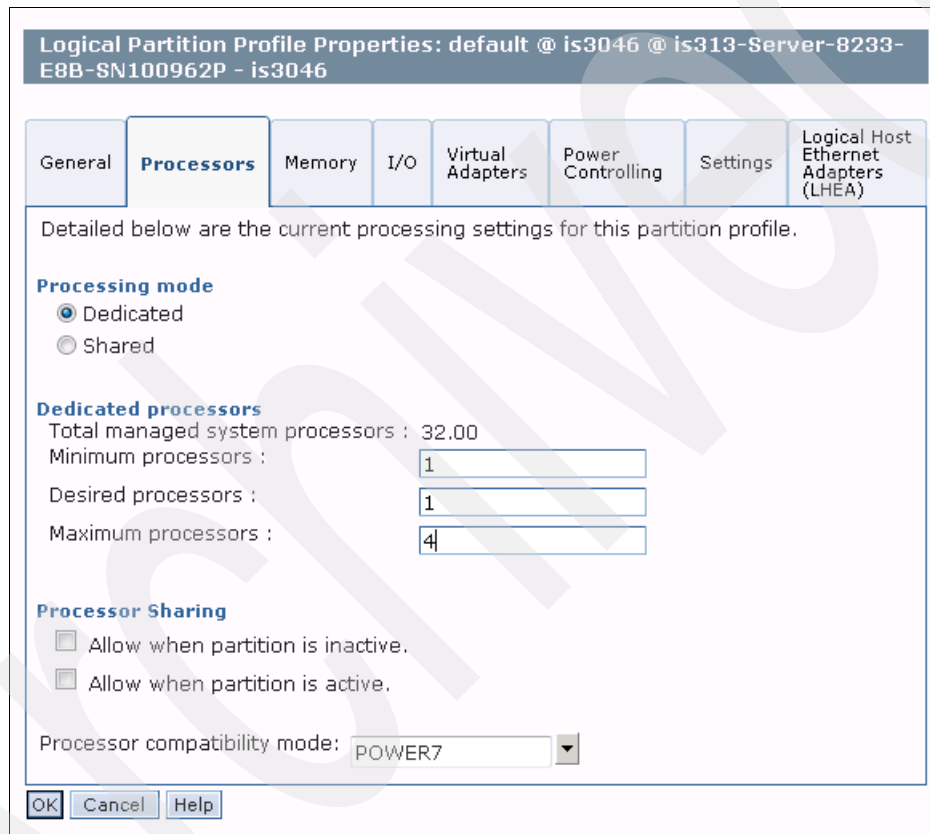


Figure 4-2 One click plus adjusting the settings to change a dedicated LPAR into a SPLPAR

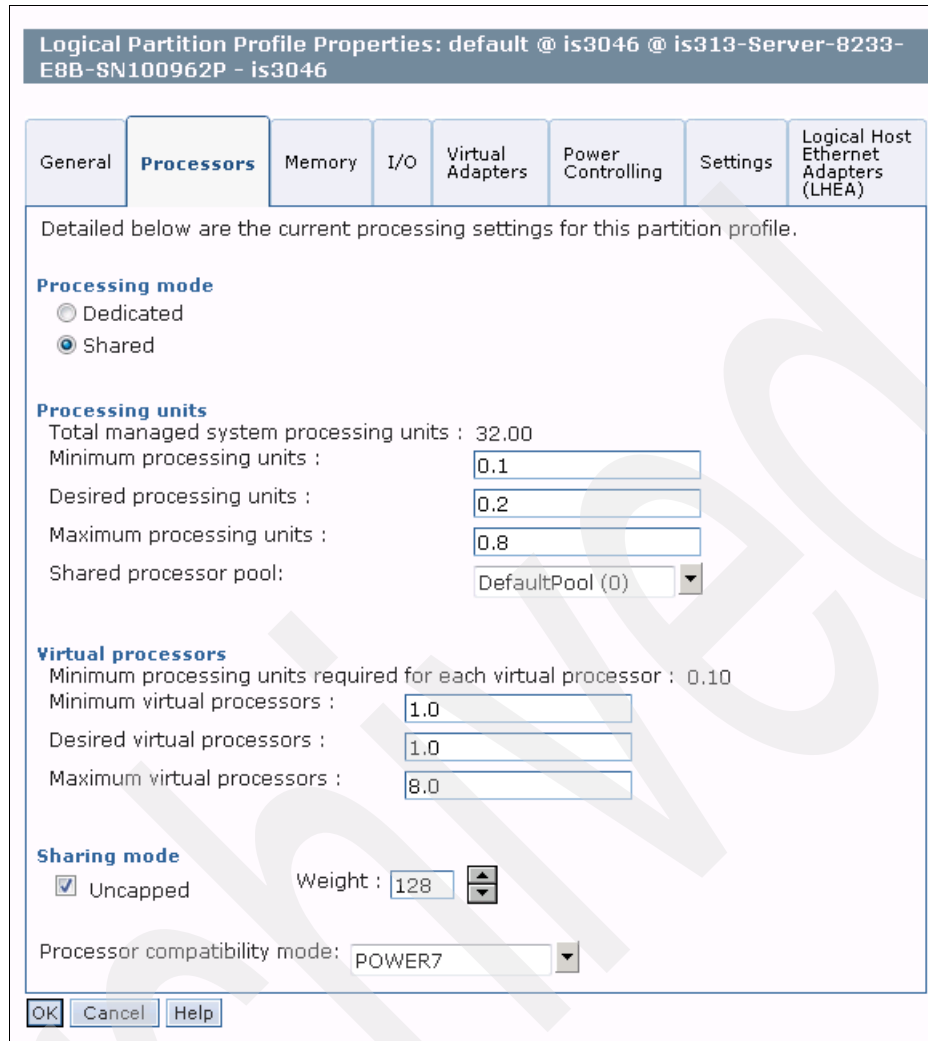


Figure 4-3 One click plus adjusting the settings to change a dedicated LPAR into a SPLPAR

## 4.2 Micropartition design option

The shared processor LPARs have two main flavors: those with strict resource definitions and those with flexible limits.

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

A shared LPAR that you define in this manner resembles a dedicated LPAR, with two exceptions: Other LPARs can use its resources when not used by their owner (more flexible than with dedicated shared partitions), and a fraction of processors can be assigned whereas Dedicated LPARS always require whole processors.

LPAR entitlements are in a sense reserved capacity: The sum of the entitlements of all of the SPLPARs sharing the pool cannot exceed the number of physical processors in the pool,

which is the only way that the entitlements are guaranteed. Therefore, large LPAR entitlements limit the number of LPARs that can be active in the shared processor pool.

To achieve the best processor sharing, keep the entitlements small and the partitions uncapped. Uncapped partitions can utilize resources in the shared pool well beyond their guaranteed entitlement, which is the most common approach for combined SAP landscapes because the LPARs can adapt to the workload requirements, and the resource distribution can adapt to the various workload peaks in multiple LPARs. Of course this also leads to contention in the shared processor pool if workloads are heavily competing for resources. You have several options to enforce a priority policy and relative behavior of the LPARs, which are referred to as implicit capping. They provide a means to control the resource allocation, while still providing the flexibility and efficiency of resource sharing.

### 4.3 Assigning implicit capping to LPARs

Most LPAR settings can be dynamically altered during runtime by DLPAR (dynamic LPAR) operations using the HMC. Using this functionality, you can manually change the behavior of the LPARs or schedule DLPAR operations using the HMC directly or by external tools.

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

By reducing the LPAR weight, you do not actually cap the LPAR; instead, you reduce its ability to compete for additional resources above its actual entitlement. If there is no competition for these resources, the LPAR is only restricted by its number of configured VCPUs, which allows it to be both flexible and harmless to other workloads of higher priority (with greater weights).

The advantage of implicit capping versus explicit capping is:

- ▶ The flexibility that is gained by low reserved entitlement, which allows more SPLPARs to be defined on the same number of physical processors on the one hand and on the other hand a second “implicit entitlement” can be guaranteed.
- ▶ The protection of SPLPARs with low weights from being completely cannibalized by uncapped SPLPAR with higher weights.
- ▶ Limit the maximum processors in hosting environments.
- ▶ Can be scheduled as described in “Scheduling dynamic LPAR operations” on page 34, which is not possible with the weight.

Figure 4-4 on page 34 shows how to access the HMC panel to dynamically change the resources.

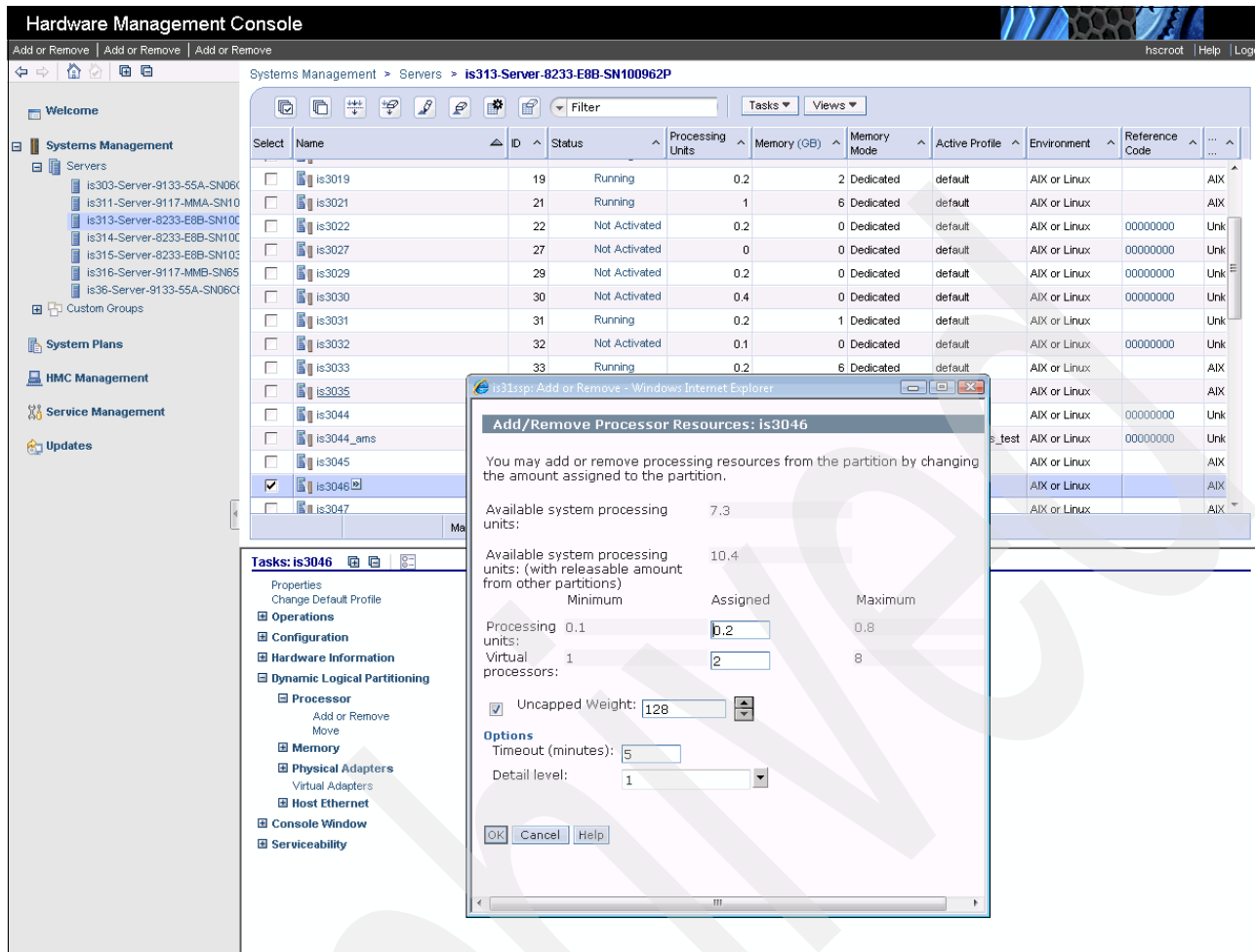


Figure 4-4 Change amount of virtual processors

**Note:** A partition that is capped and a partition that is uncapped with a weight of 0 are functionally identical. In terms of performance, you get exactly the same result (utilization can only go up to the partition's entitled capacity, not higher). In an SAP system, this has an impact in the monitoring because a partition is visible as uncapped, yet behaves as a capped partition. See: <http://www.redbooks.ibm.com/abstracts/tips0538.html>

## 4.4 Scheduling dynamic LPAR operations

The HMC provides the possibility to schedule dynamic LPAR operations to add or remove processor and memory, and assign or unassign I/O devices at a defined period in time. You can use this scheduling to change the LPAR resource allocation, for example, to support an SAP mode change.

Earlier in this book, we discussed the advantages provided by the dynamic reconfiguration feature. In this example, we focus on dynamically changing resource allocation of the processor. To use this functionality for a specific LPAR:

1. In the left frame of the HMC, select **Systems Management** → **Servers**, and choose the server that the LPAR belongs to.

2. Choose the LPAR to which you want to apply the operation. Select **Operations** → **Schedule operations** to get to Figure 4-5.



Figure 4-5 HMC menu selection

3. To allocate resources during runtime, select **Dynamic Reconfiguration**, as shown in Figure 4-6.

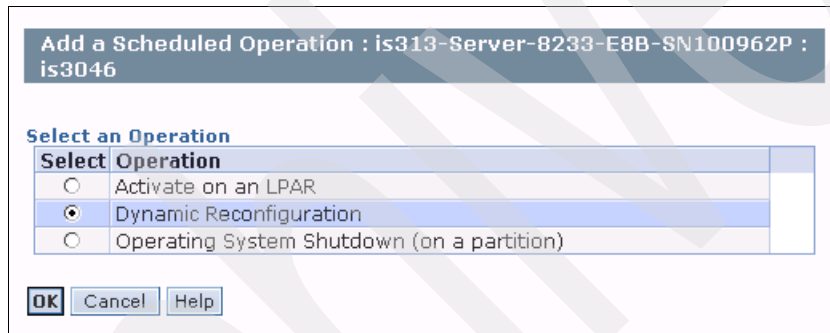


Figure 4-6 HMC menu selection

- To add virtual processors on top of the already assigned ones, select the **Options** tab as shown in Figure 4-7.

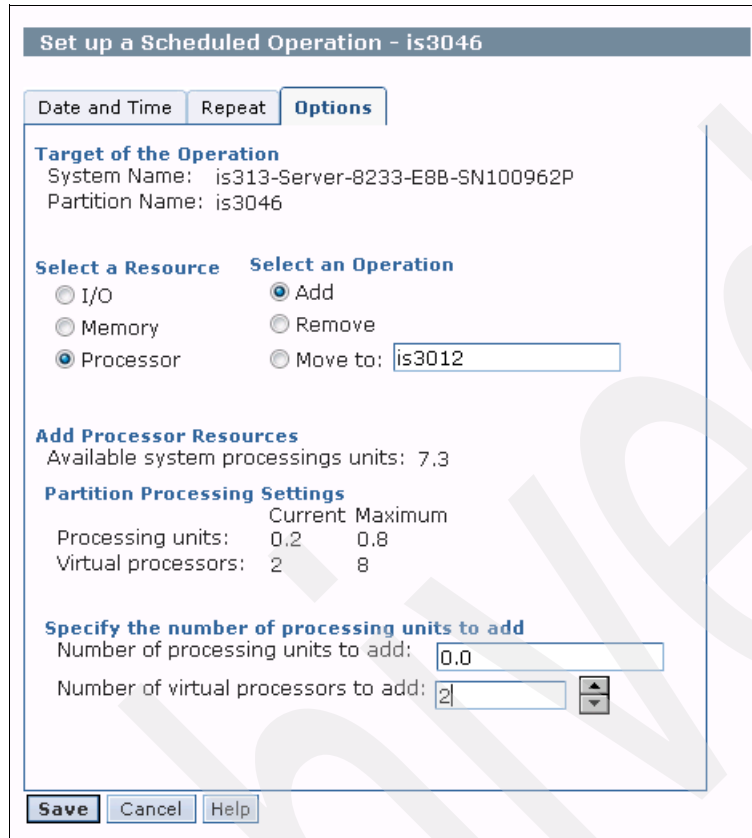


Figure 4-7 HMC menu selection

- Perform the scheduling by clicking the **Date and Time** tab, which was already visible in Figure 4-7. To repeat this task, select the **Repeat** tab to define how often you want the event to repeat. Save and exit the configuration.

This scheduling allows the LPARs to react on predictable load peaks. To give a small processor-based example, we look into a machine with an LPAR per time zone having predictable peaks during the day and low utilization during the night. Hence they have different load peaks moving from partition to partition. By changing the amount of virtual processors, we can ensure that batch processes during the night shift of an LPAR do not impact the performance of a partition running in day shift and better utilize the processor resources of the shared pool.

## 4.5 Managing multiple shared processor pools

Multiple shared processor pools (MSPP) are useful in SAP landscapes in several ways, including the ability to segregate workloads and the ability to create subpools, which you can use to limit software license costs that are based on the number of available processors.

There are several IBM Redbooks publications available that describe in detail how you manage multiple shared processor pools. This section is an extract of the IBM Redpaper:

*IBM System p Advanced POWER Virtualization (PowerVM) Best Practices*, REDP-4194

<http://www.redbooks.ibm.com/abstracts/redp4194.html?Open>

Starting with the POWER6 technology-based systems, you can define multiple shared processor pools and assign the shared partitions to any of the MSPPs. Setting up this feature is rather simple because you only set the properties of a processor pool.

To set up a shared processor pool (SPP):

1. Select the system on the HMC, and select **Configuration** → **Shared Processor Pool Management**. Set up the shared processor pool.
2. Assign a partition to the pool that you just set up using the same interface. Click the **Partitions** tab, and select the partition name, as shown in Figure 4-8.

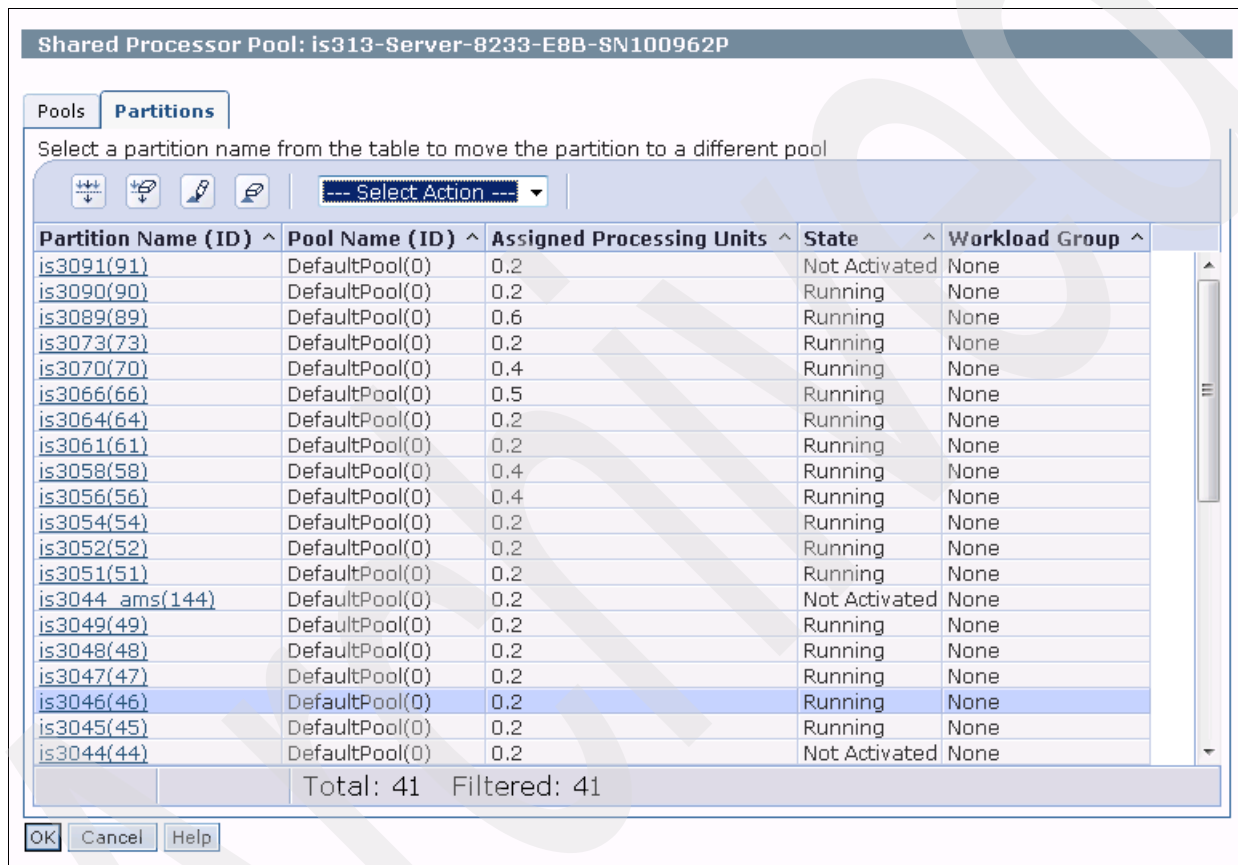


Figure 4-8 Partition assignment to multiple shared processor pools

3. Select the SPP that you want to assign this partition to, as shown in Figure 4-9.

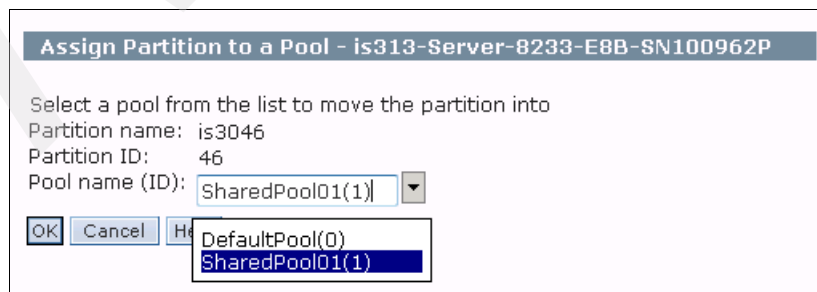


Figure 4-9 Assign a partition to a shared processor pool

## 4.6 Enable Pool Utilization Authority

In a shared pool environment the Pool Utilization Authority (PUA) can be used to restrict the data, which can be monitored within an LPAR. This can be necessary for example in hosted environments, where customers should not be able to get information about the hosting system. On the other side monitoring of processor usage in SAP may require additional metrics from the processor pool. The table in 9.3, "Other monitoring tools" on page 117 shows which metrics are affected.

Unfortunately the HMC does not provide on any screen a setting explicitly named PUA or Pool Utilization Authority. Therefore we show in this short section how to change this setting.

Select in the HMC an LPAR and open the properties as in Figure 4-10:

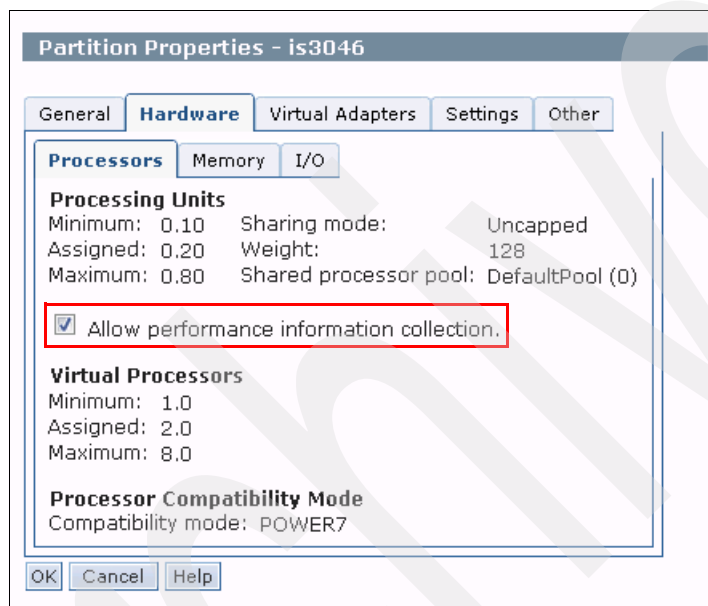


Figure 4-10 HMC LPAR Properties

On the Hardware Tab chose the page for Processors. PUA is enabled by setting the checkbox for "Allow performance information collection".

## 4.7 Activate and modify memory compression

Starting with POWER7 technology-based systems and AIX 6.1 memory compression can be used and is named Active Memory Expansion (AME).



More details can be found in 8.5, “Active Memory Expansion for SAP systems” on page 84. AME can be enabled and configured on the HMC by setting the checkbox for Active Memory Expansion in the profile for an LPAR on the memory tab (see Figure 4-11 below):

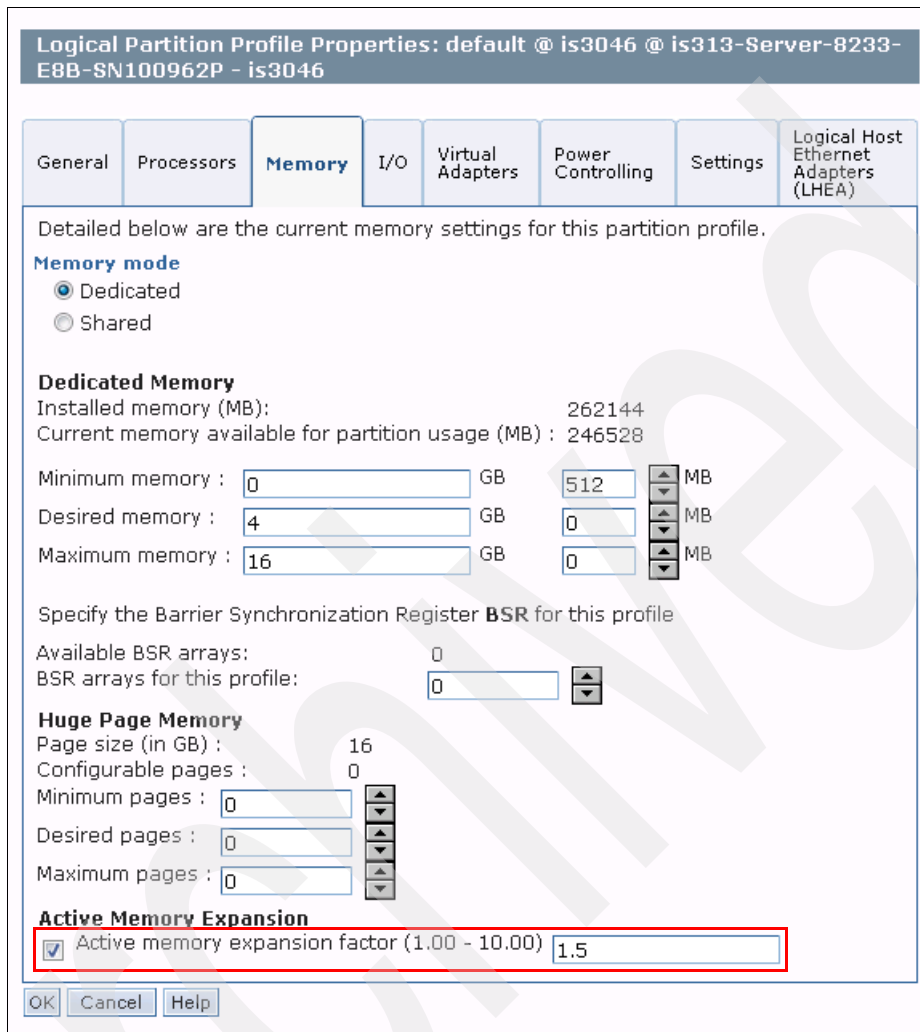


Figure 4-11 HMC LPAR Properties

On this tab you can set also the expansion factor. As you see in the example above the expansion factor of 1.5 leads to 6GB of expanded and 4GB of physical memory when the LPAR is started with this profile.

Switching AME On or Off requires a restart of the LPAR. On the other side the expansion factor can be modified without rebooting the LPAR by using DLPAR operations. On the HMC select in the LPAR list, Dynamic Logical Partinioning 'Memory' Add or Remove and change the value for the memory expansion factor (see Figure 4-12).

**Add/Remove Memory Resources - is3046**

You may add or remove memory from the partition by specifying the amount of memory the partition should have by changing the memory assigned to the partition.

	Gigabytes	Megabytes
Available system memory:	16	768
Available system memory: (with releasable amount from other partitions)	22	768
Minimum memory:	0	512
Maximum memory:	16	0
Assigned memory:	4	0

**Active Memory Expansion**  
Active memory expansion factor (1.00 - 10.00)

**Options**  
Timeout (minutes) :   
Detail level :

Figure 4-12 HMC: Add / Remove memory resources

When AME is enabled, the operating system is switching off the support for 64K pages. If necessary you can turn on 64K memory pages again by changing the restricted `vmo` setting `vmm_mpsize_support` to 1. You can do this with the following command:

```
# vmo -F -r -o vmm_mpsize_support=1
```

**Note:** Currently it is not recommended to use 64K memory pages with AME since performance impacts may be observed, when higher decompression rates are expected.

## 4.8 Simultaneous Multithreading and compatibility modes

Simultaneous Multithreading (SMT) was introduced with POWER5 processors and started as two-way multithreading (SMT-2). With the POWER7 processor, four-way multithreading (SMT-4) was introduced. In addition to the POWER7 technology-based hardware, one of the following software levels is required as the minimum level to support SMT-4:

- ▶ AIX 6.1 TL4
- ▶ IBM i 6.1.1
- ▶ Linux SLES 11 or RHEL6

With operating system commands, you can switch Simultaneous Multithreading on or off, or you can select SMT-1 (= off), SMT-2, or SMT-4 directly. Changes to the SMT mode affect the

whole partition. On AIX and Linux, changes to the SMT mode take effect immediately, on IBM i, the changes will take effect after an IPL of the partition. The following commands are used in the supported operating systems:

## AIX

You can control the SMT mode with the `smtctl` command. To switch on SMT, execute the following command:

```
# smtctl -m on
```

To switch off SMT, use the command:

```
# smtctl -m off
```

In order to switch between the three SMT modes on a POWER7 technology-based machine, the command `smtctl` has been enhanced with another parameter to change the number of SMT threads. For example, in order to run the LPAR with four SMT threads you execute the following command:

```
# smtctl -t 4
```

## IBM i

You can control the SMT mode through system value `QPRCMLTTSK`. There are three possible values for that system value: Value 0 to turn SMT off, value 1 to turn SMT on to the highest available degree (SMT-2 on POWER6, SMT-4 on POWER7), and value 2 to let the operating system decide and dynamically change what level of SMT to use. Selecting SMT-2 explicitly on POWER7 technology-based systems is currently not supported on IBM i. To switch on SMT, enter the command:

```
CHGSYSVAL SYSVAL(QPRCMLTTSK) VALUE('1')
```

To switch off SMT, enter the command:

```
CHGSYSVAL SYSVAL(QPRCMLTTSK) VALUE('0')
```

## Linux

You can control the SMT mode with the `ppc64_cpu` command with parameter `--smt`. You can select either on or off, or you can select the SMT mode directly. To switch on SMT, execute the command:

```
ppc64_cpu --smt=on
```

or:

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

To switch off SMT, execute the command:

```
ppc64_cpu --smt=off
```

or:

```
ppc64_cpu --smt=1
```

To select SMT-2 explicitly, execute the command:

```
ppc64_cpu --smt=2
```

To display the current SMT status, execute the command:

```
ppc64_cpu --smt
```

With the Hardware Management Console (HMC), you can define a compatibility mode in the LPAR profile. The compatibility mode allows LPAR migration between Power Servers running with different versions of POWER processors. Currently supported are the compatibility modes POWER6, POWER6+™, and POWER7. If you are running on POWER7 technology-based systems and switch the compatibility mode in the profile to POWER6 or POWER6+, you implicitly turn off the capability to run with four SMT threads. The following Figure 4-13 is showing where to change the compatibility mode in an LPAR profile on the HMC:

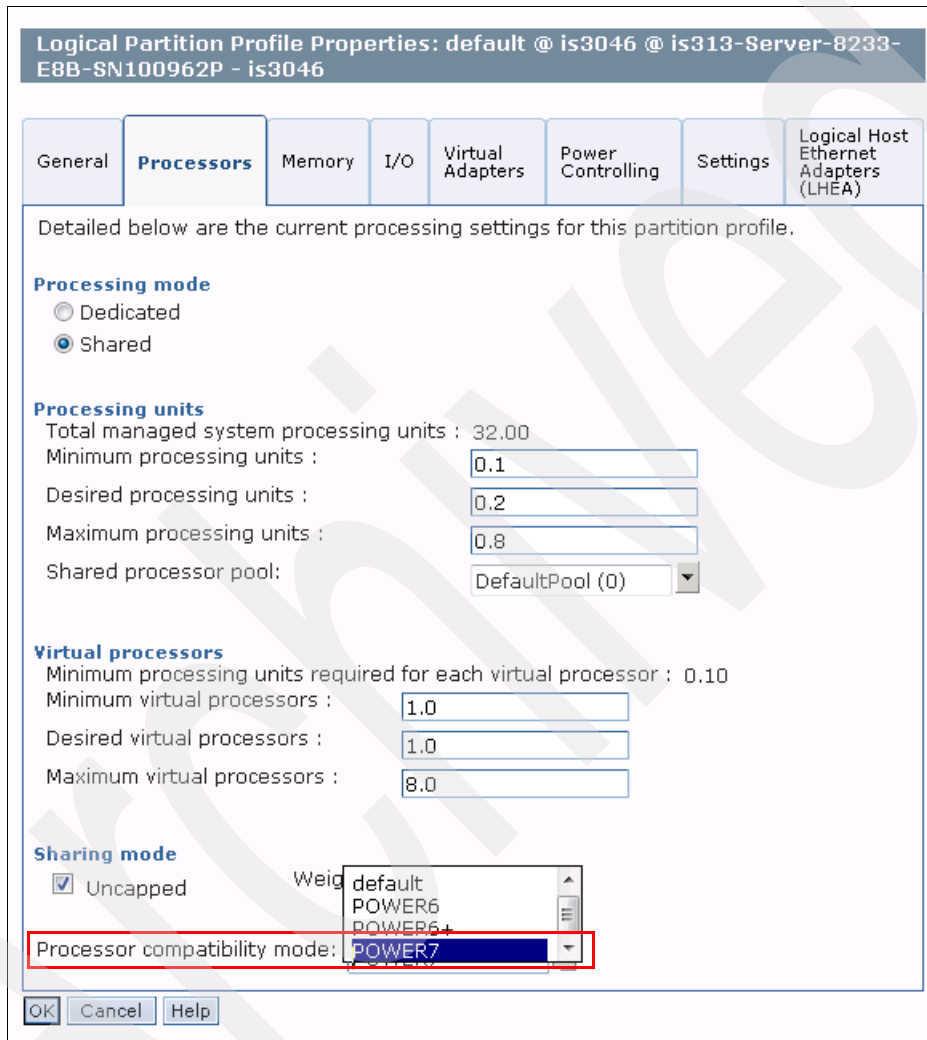


Figure 4-13 Selecting Processor compatibility mode in the HMC

**Note:** The purpose of the compatibility mode is to allow LPAR migration. You should not use the compatibility mode when your only intention is to force SMT-2 on POWER7 technology-based hardware. To switch between SMT-2 and SMT-4, use the previously described operating system commands.

## Virtual I/O Server

SAP clients have used the Virtual I/O Server (VIOS) in their IT landscapes for several years. Multiple white papers and technical reference papers are available to document their experiences.

The white paper *System landscape at Stadtwerke Düsseldorf on a virtualized IBM POWER5 environment*, GM 12-6851-00, describes the system landscape of a large utility provider in Germany. It covers the technical implementation and some utilization data of the VIOS, as shown in Figure 5-1.

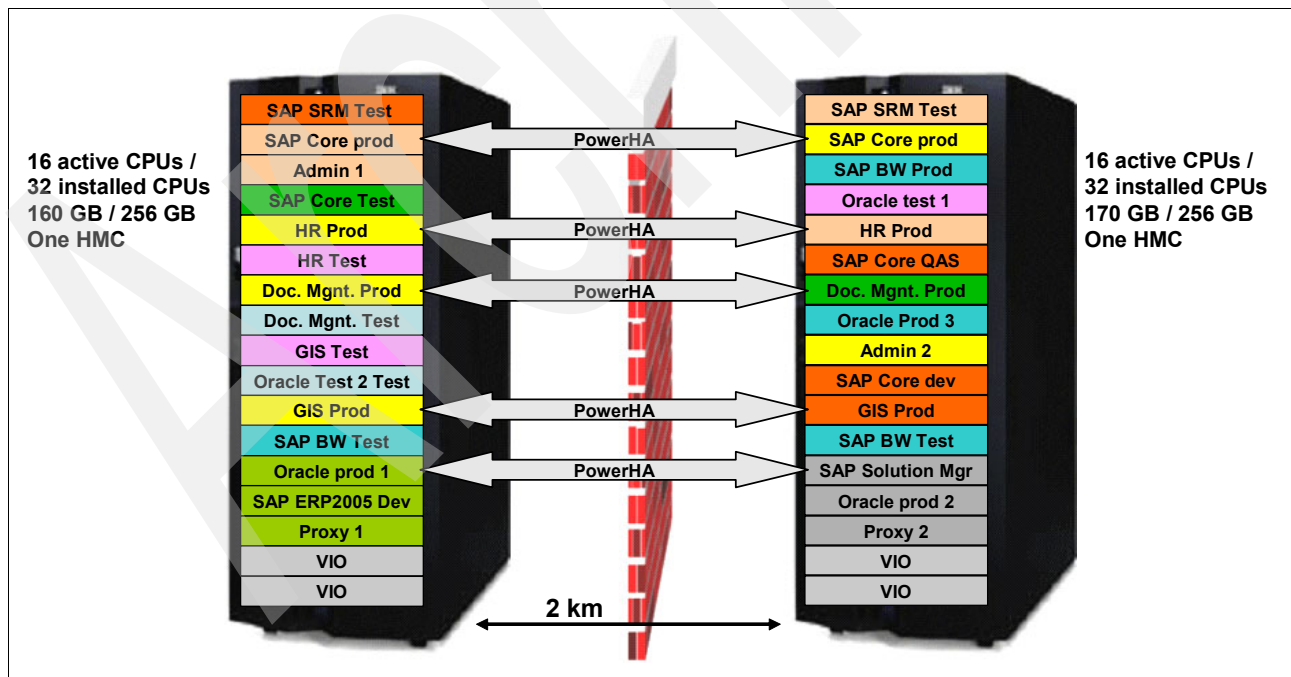


Figure 5-1 IT landscape of Stadtwerke Düsseldorf from 2007

A large-scale example of an automotive supplier is described in the technical reference paper *Advantages of IBM POWER technology for mission-critical environments*, GE 12-2885, which is available at:

[http://www.ibm.com/software/success/cssdb.nsf/CS/STRD-79YNA4?OpenDocument&Site=default&cty=en\\_us](http://www.ibm.com/software/success/cssdb.nsf/CS/STRD-79YNA4?OpenDocument&Site=default&cty=en_us)

## 5.1 Motivation

Using the Virtual I/O Server you can define logical partitions (LPARs) independent from physical I/O resources. The VIOS itself runs in a special logical partition. The physical I/O resources, such as SCSI controllers and network adapters, are assigned to the VIOS partition. The VIOS provides virtual SCSI adapters, virtual network and virtual Fibre Channel access to the LPARs and maps the virtual resources to the physical resources. The LPARs only use the virtual network, virtual SCSI, or Fibre Channel adapters that the VIOS provides.

Using a VIOS provides the following advantages:

- ▶ You can use LPARs independently from physical I/O resources. Several LPARs can share the same physical I/O resource; therefore, the amount of physical adapters is reduced.
- ▶ Because all physical resources are assigned to the VIOS, no special hardware drivers are needed inside the client LPARs. The LPARs only need drivers for a virtual SCSI (vSCSI) host adapter and a virtual network adapter. These drivers are delivered together with the operating system.
- ▶ On Linux LPARs, using a VIOS avoids using proprietary hardware drivers in the Linux operating system.
- ▶ Using a VIOS is a prerequisite for using the Live Partition Mobility feature. See also Chapter 6, “IBM PowerVM Live Partition Mobility” on page 63.

In the first part of this chapter, we describe the basic usage types of a VIOS and show the schematic setup for using virtual network and virtual SCSI.

In the next part of this chapter, we show the management tasks that are necessary to set up a VIOS and connect a client LPAR to it.

In the last part of this chapter, we describe how to monitor a VIOS.

More detailed information and advanced scenarios are described in the IBM Redpaper publication *Advanced POWER Virtualization on IBM System p Virtual I/O Server Deployment Examples*, REDP-4224.

## 5.2 Virtual I/O Server: basic usage types

Next we cover:

- ▶ Using virtual SCSI
- ▶ High availability for vSCSI
- ▶ Using virtual network
- ▶ High availability for virtual networks

## 5.2.1 Using virtual SCSI

The VIOS provides disk resources to LPARs using virtual SCSI (vSCSI) server adapters. Inside the VIOS, the administrator maps a disk resource (backing device) to a vSCSI server adapter. The backing device can be a physical hard disk or a logical volume inside a volume group. A vSCSI target device provides the connection between the vSCSI server adapter and the backing device.

The client LPAR uses a vSCSI client adapter to connect to the disk resource through the VIOS. Figure 5-2 shows a vSCSI setup with three client LPARs connected to the VIOS. Inside the operating system of the client LPAR, the hard disk resource is recognized as a standard SCSI disk device (for example, `hdisk0` on AIX, `DC01` on IBM i and `/dev/sda` on Linux).

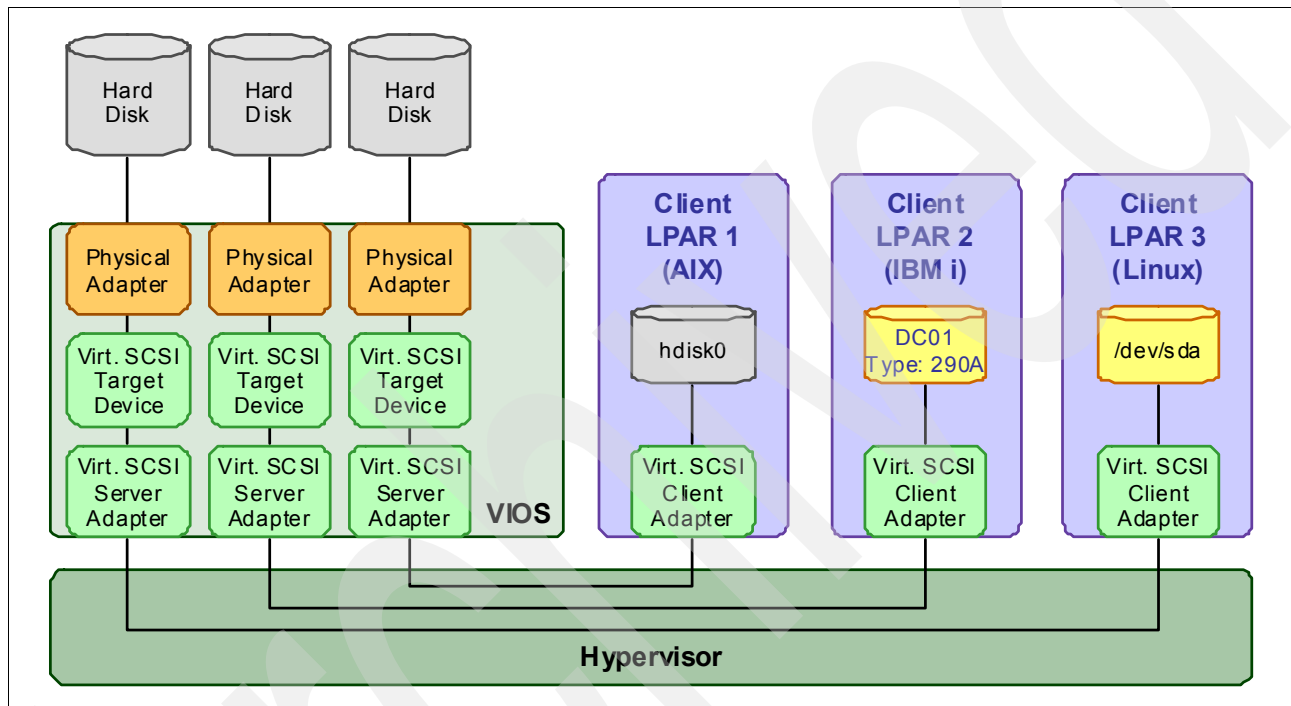


Figure 5-2 Basic virtual SCSI setup

## 5.2.2 High availability for vSCSI

The setup in Figure 5-2 does not provide any high-availability features. If the VIOS fails, all client LPARs connected to it are disconnected from the hard disk resources.

To achieve a higher availability, you can set up two redundant Virtual I/O Servers. Both are connected to the same hard disk resource (for example, a SAN volume). Each VIO Server provides this resource using vSCSI to the client partition. The client partition uses multipath I/O (MPIO) features to connect to this resource. If one VIOS fails, the hard disk access is still possible through the second VIOS. Figure 5-3 on page 46 shows the schematic setup for this usage type.

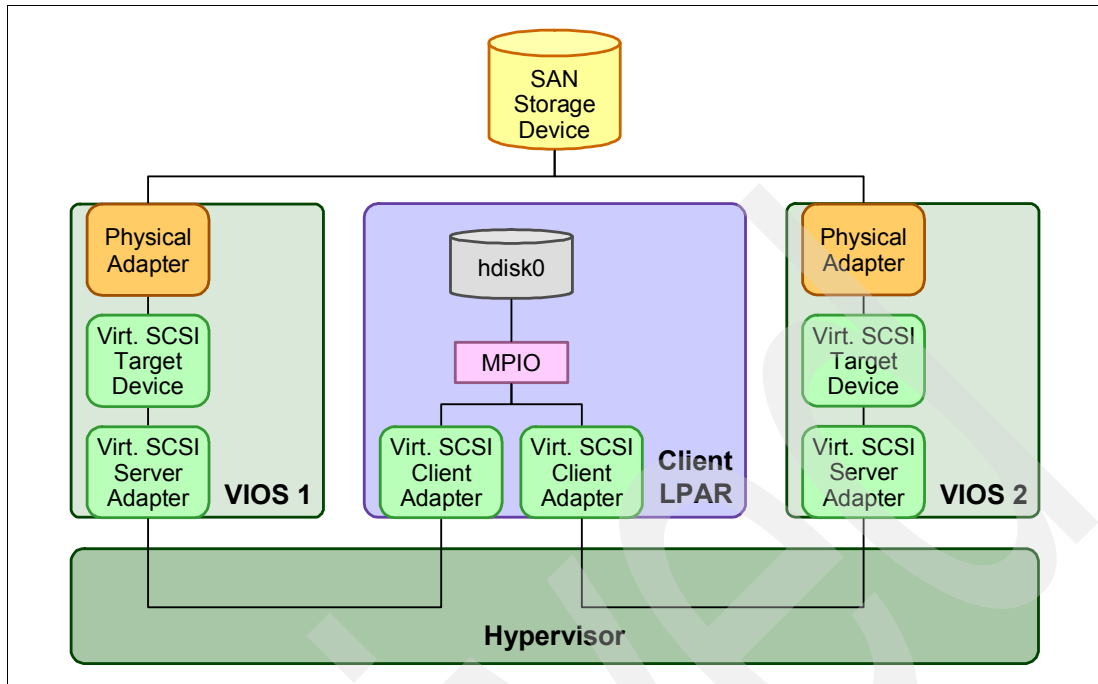


Figure 5-3 Redundant access to hard disk resources using two VIO servers

### 5.2.3 Using the virtual network

The POWER Hypervisor implements a virtual LAN (VLAN) aware Ethernet switch that the client LPARs and the Virtual I/O Server can use to provide network access. Figure 5-4 shows a scenario using the virtual network.

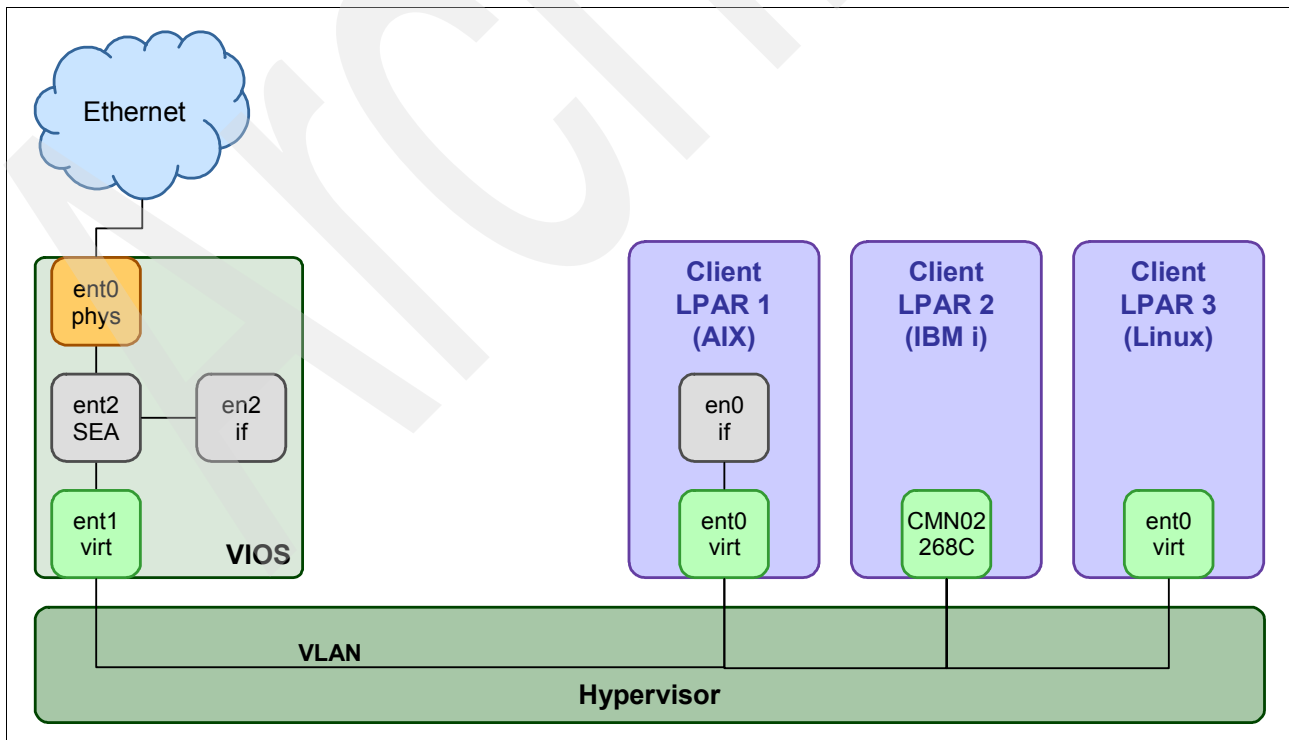


Figure 5-4 Basic virtual network setup



The physical network interfaces are assigned to the VIOS partition. In Figure 5-4 on page 46, ent0 is the physical adapter that is assigned to the VIOS.

Virtual Ethernet adapters are assigned to the VIOS and the client LPAR. All virtual Ethernet adapters that use the same virtual LAN ID are connected through a virtual Ethernet switch that the Hypervisor provides. In Figure 5-4 on page 46, ent1 is the virtual network adapter of the VIOS. In the client LPARs, the virtual network adapters appear as standard network interfaces (for example, ent0 in the AIX LPAR, CMN02 in the IBM i LPAR, and eth0 in the Linux LPAR).

The VIOS uses a Shared Ethernet Adapter (SEA) to connect the virtual Ethernet adapter to the physical adapter. The SEA acts as a layer-2 bridge, and it transfers packets from the virtual Ethernet adapter to the physical adapter. In Figure 5-4 on page 46, ent2 is the shared Ethernet adapter, and en2 is the corresponding network interface.

### 5.2.4 High availability for virtual networks

The setup in Figure 5-4 on page 46 does not provide any high-availability features. If the VIOS fails, all connected LPARs lose their network connection. To overcome this single point-of-failure, use a setup with two Virtual I/O Servers. Each VIOS uses a separate VLAN ID to provide network access to the client LPARs. The client LPARs use link aggregation devices (LA) for the network access. If one VIOS fails, network access is still possible through the remaining VIOS. Figure 5-5 shows the schematic setup for this usage type.

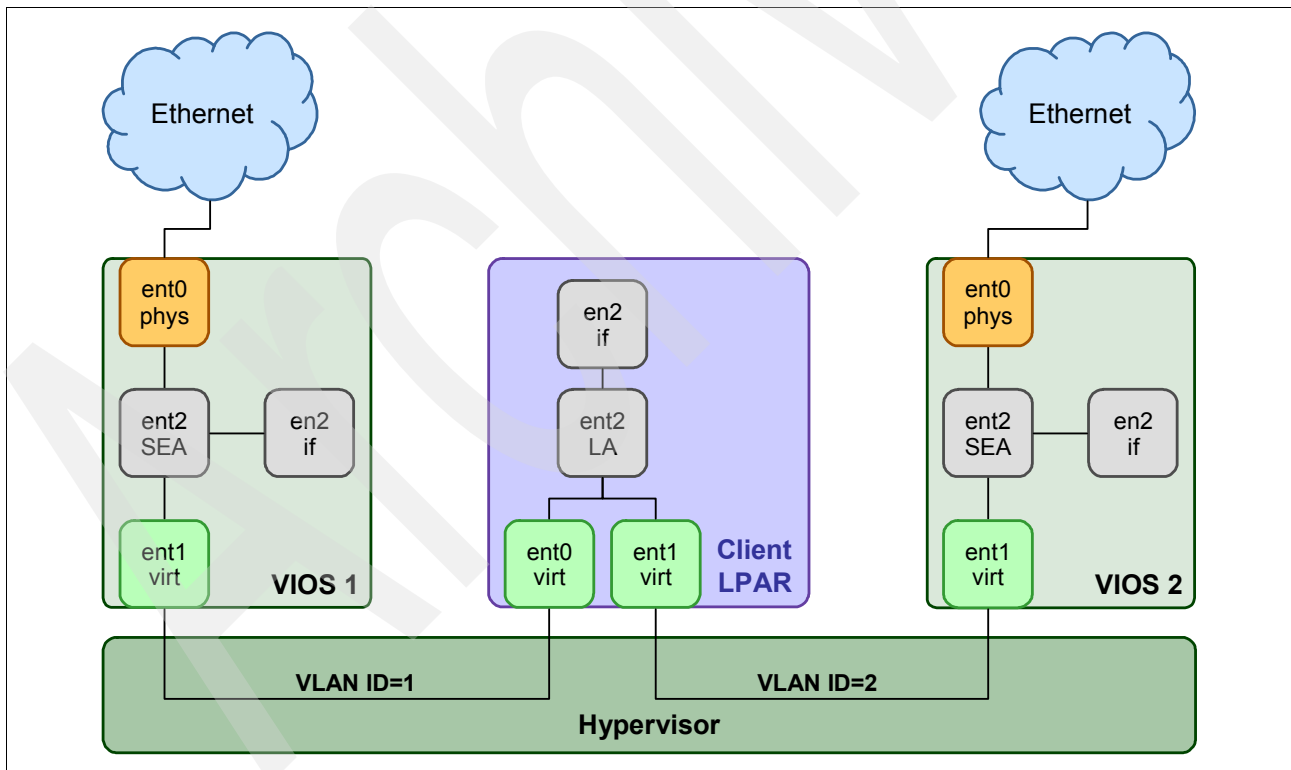


Figure 5-5 Redundant network access using two Virtual I/O Servers

## 5.3 Setting up a VIOS partition

Next we discuss:

- ▶ Defining the VIOS LPAR
- ▶ Installing the VIOS
- ▶ Creating virtual SCSI server adapters
- ▶ Gathering information about existing virtual adapters
- ▶ Connecting a client LPAR to a virtual SCSI server adapter
- ▶ Creating virtual Ethernet adapters
- ▶ Connecting a client LPAR to the virtual network
- ▶ TCP/IP address for the VIOS

### 5.3.1 Defining the VIOS LPAR

To set up a new VIOS, create a new LPAR and use the partition type “VIO Server”. It is recommended that you create an uncapped shared processor partition and assign all required physical I/O resources to it. The amount of RAM that the VIOS needs does not depend on the number of client partitions or on the amount of expected I/O operations. Assigning one GB of RAM to the VIOS partition is a good choice. Run the VIOS partition in shared processor mode. Defining two virtual processors is sufficient for most cases.

**Important:** The VIOS partition must have the highest uncapped weight on the system. If there are two VIOS defined, give each of them the same weight.

### 5.3.2 Installing the VIOS

The VIOS is based on the AIX operating system. The installation is the same as a standard AIX operating system installation, but uses a different installation source:

1. After the installation finishes, log on as user *padmin*, and define a password for this user.
2. Before you use the VIOS, you must accept the VIOS license using the `license -accept` command.

### 5.3.3 Creating virtual SCSI Server Adapters

To create a virtual SCSI server adapter:

1. In the partition profile on the HMC, click the **Virtual Adapters** tab, and select **Action** → **Create** → **SCSI Adapter**, as shown in Figure 5-6 on page 49 and Figure 5-7 on page 49.

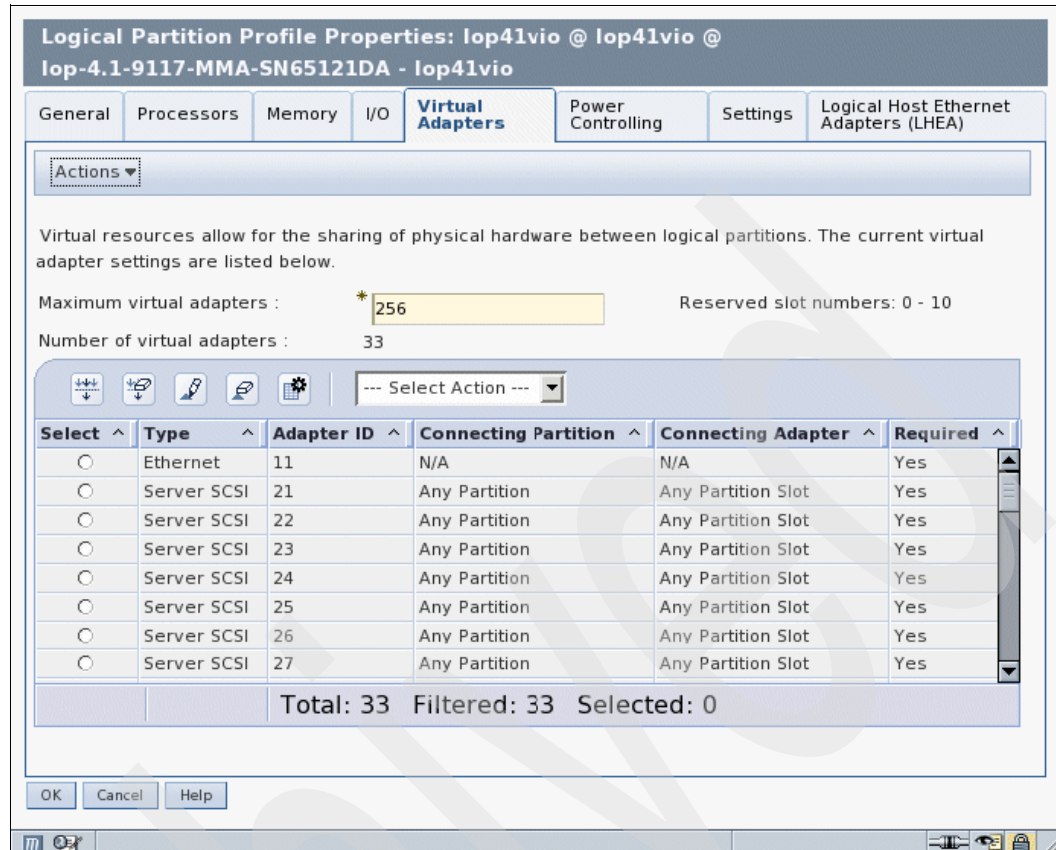


Figure 5-6 Virtual SCSI server adapters in the VIOS LPAR profile

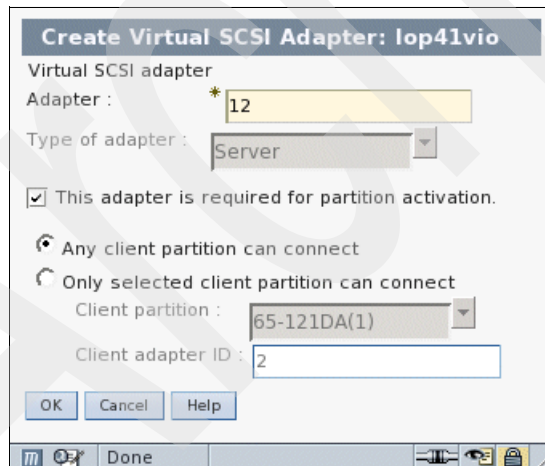


Figure 5-7 Create a Virtual SCSI server adapter

In the VIOS, the virtual SCSI server adapters appear as vhost-devices:

```
$ lsdev -virtual
name          status      description
[...]
vhost0       Available  Virtual SCSI Server Adapter
vhost1       Available  Virtual SCSI Server Adapter
vhost2       Available  Virtual SCSI Server Adapter
```

vhost3 Available Virtual SCSI Server Adapter

2. After you set up the virtual SCSI server adapter, map it to the physical resource. To map the virtual adapter to the physical resource, create a mapping device using the `mkvdev` command:

```
$ mkvdev -vdev hdisk28 -vadapter vhost0 -dev vls1234_sda
```

In this example, the backing device `hdisk28` is mapped to the virtual SCSI server adapter `vhost0`. The name of the mapping device is `vls1234_sda`.

### 5.3.4 Gathering information about existing virtual adapters

Sometimes it is useful to obtain the actual configuration of a virtual SCSI server adapter. To do so, execute the `lsmmap` command, as shown in Example 5-1.

*Example 5-1 Obtaining configuration of a virtual SCSI server adapter*

```
$ lsmmap -vadapter vhost0
SVSA          Physloc          Client Partition ID
-----
vhost0       U9117.MMA.65121DA-V2-C21  0x00000004

VTD          vls1234_sda
Status       Available
LUN          0x8100000000000000
Backing device hdisk28
Physloc      U789D.001.DQD51K5-P1-C1-T1-W200300A0B817B55D-L4000000000000
```

To get information about all existing virtual SCSI server adapters, run the `lsmmap -a11` command.

### 5.3.5 Connecting a client LPAR to a virtual SCSI server adapter

To connect a client LPAR to a hard disk resource, you must define a virtual SCSI client adapter in the partition profile of the client LPAR:

1. Open the partition profile on the HMC, and click the **Virtual Adapters** tab. Select **Actions** → **Create** → **SCSI adapter**. Choose an adapter ID, and mark the adapter as a required resource for partition activation, as shown in Figure 5-8.

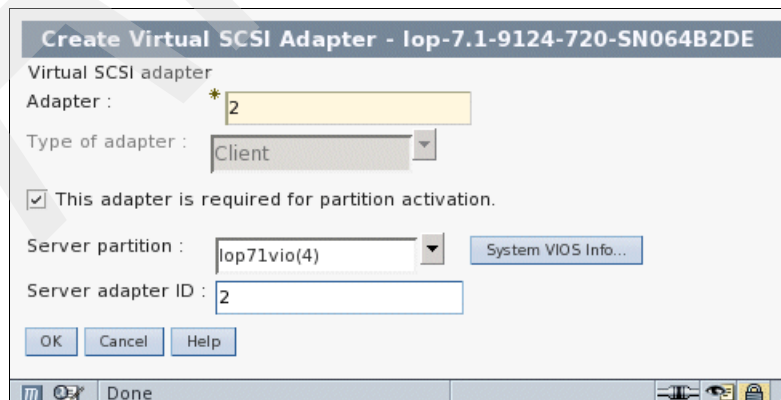


Figure 5-8 Create vSCSI Client Adapter

2. Press **System VIOS Info** to get an overview of vSCSI server adapters on the selected VIOS partition. Figure 5-9 shows an overview of server adapters.
3. Select the required resource and activate the partition.

**Virtual I/O Server Information - lop-7.1-9124-720-SN064B2DE**

The Virtual I/O Servers and their virtual scsi devices are detailed below. Adapters that have a status of 'Active' are currently hosting a client virtual scsi adapter.

You may specify the client adapter's corresponding server adapter by selecting a server adapter from the choices below.

Select	Server	Server Slot	Status	Virtual Adapter	Backing Device	Client Partition	Client Slot	Client Disks
<input type="radio"/>	lop71vio(4)	38	Inactive	vhost8		Any Partition	Any Partition Slot	
<input type="radio"/>	lop71vio(4)	37	Inactive	vhost7		Any Partition	Any Partition Slot	
<input type="radio"/>	lop71vio(4)	39	Inactive	vhost9		Any Partition	Any Partition Slot	
<input type="radio"/>	lop71vio(4)	34	Inactive	vhost4		Any Partition	Any Partition Slot	
<input type="radio"/>	lop71vio(4)	33	Active	vhost3	hdisk7	ls3742(5)	33	
<input type="radio"/>	lop71vio(4)	36	Inactive	vhost6		Any Partition	Any Partition Slot	
<input type="radio"/>	lop71vio(4)	35	Inactive	vhost5		Any Partition	Any Partition Slot	
<input type="radio"/>	lop71vio(4)	31	Active	vhost1	hdisk5	ls3767(3)	31	
<input type="radio"/>	lop71vio(4)	32	Active	vhost2	hdisk6	ls3743(6)	32	
<input type="radio"/>	lop71vio(4)	30	Active	vhost0	hdisk4	ls3766(2)	30	

OK Cancel

Done

Figure 5-9 Overview of vSCSI server adapters

### 5.3.6 Creating virtual Ethernet adapters

Virtual Ethernet adapters are defined in the LPAR profile on the HMC:

1. In the VIOS partition profile, click the **Virtual Adapters** tab, and select **Actions** → **Create** → **Ethernet Adapter**. Select the **This adapter is required for partition activation** option. Because this adapter is used to connect the client LPARs to the external network, select the **Access external network** option, as shown in Figure 5-10.

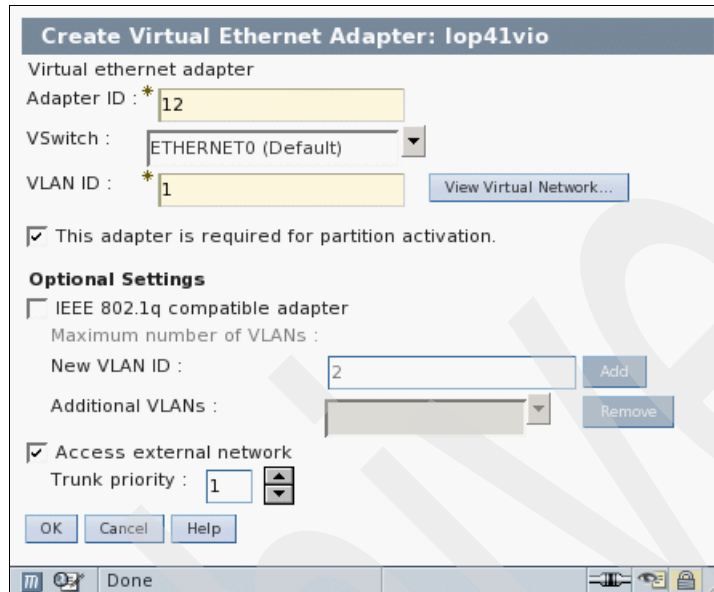


Figure 5-10 Create a virtual Ethernet adapter in the VIOS partition profile

After you activate the changes, the new adapter appears in the VIOS:

```
$ lsdev -virtual
name          status      description
[...]
ent1          Available  Virtual I/O Ethernet Adapter (1-lan)
[...]
```

2. Create the Shared Ethernet Adapter (SEA) to connect the virtual adapter to the physical adapter:

```
$ mkvdev -sea ent0 -vadapter ent1 -default ent1 -defaultid 1
ent2 Available
```

In Figure 5-10, ent0 is the physical adapter, and ent1 is the VLAN adapter. The parameter defaultID defines the VLAN ID to use. The new shared adapter is created with the name ent2.

### 5.3.7 Connecting a client LPAR to the virtual network

In the partition profile of the client LPAR, create a virtual Ethernet adapter that is similar to the setup of the VIOS. Do *not* select the **Access external network** option for the client LPAR because the network access is provided by the VIOS; instead, click **View Virtual Network** to get an overview of all VLAN IDs in the system.

In the client LPAR, the virtual network device appears as a standard network device (for example, ent0 in AIX and eth0 in Linux). Configure it similar to a standard physical network device.

### 5.3.8 TCP/IP address for the VIOS

The VIOS does not necessarily need a TCP/IP network address. The VIOS can be accessed from the HMC using the secure and private HMC to service processor network, as shown in Figure 5-11.

You must assign a TCP/IP address to the VIOS, at least, for the following reasons:

- ▶ Upgrading the VIOS through the network

You can upgrade the VIOS using a CD or using network access (for example, through FTP). Upgrading the VIOS using a network needs a TCP/IP address assigned to it, at least for the time of the upgrade.

- ▶ Monitoring the VIOS

If you plan to monitor the VIOS, as described in 5.4, “VIO Server monitoring” on page 54, assign a TCP/IP address to it permanently.

- ▶ Prerequisite for Live Partition Mobility

There are two possibilities of assigning a TCP/IP address to the VIOS:

- Use a free physical network adapter that is not used for the communication between the public network and the client LPARs.
- Create an additional virtual network adapter in the VIOS, and assign a TCP/IP address to it, for example, using the **mktcpip** command. You might want to put the VIOS on a specific VLAN that is different from the VLAN that the LPARs use. See Figure 5-11.

For more information, visit:

<https://www.ibm.com/developerworks/eserver/library/es-pwr5-virtualvlan/>

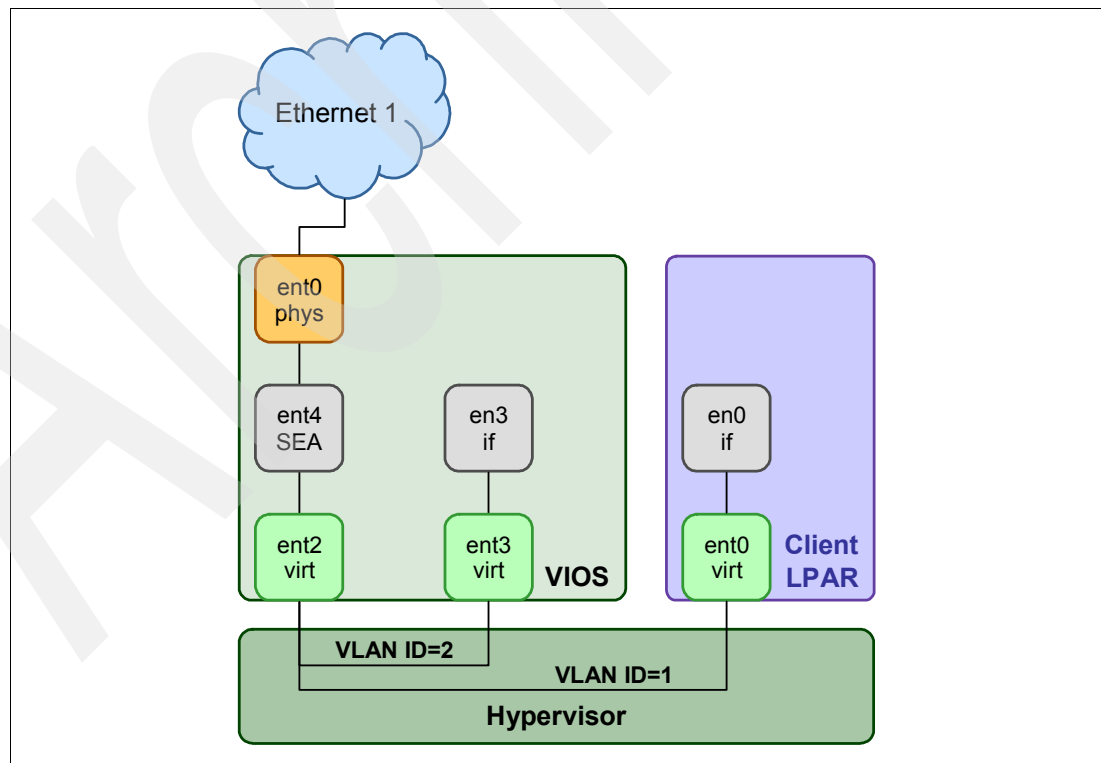


Figure 5-11 TCP/IP address for the VIOS using an additional virtual network adapter (ent3/en3)

### 5.3.9 N\_Port ID Virtualization

With a new generation of fiber channel adapters, the VIO server also supports N\_Port ID Virtualization (NPIV) as an alternative technology to attach SAN devices to VIO clients. NPIV is a fiber channel standard that enables multiple fiber channel initiators to share a single physical fiber channel port. PowerVM facilitates this feature with the introduction of virtual fiber channel adapters, which can be defined and mapped in the partition profiles for the VIO server and clients. The virtual fiber channel adapters are then connected to a physical fiber channel port.

After the initial setup the client partitions can access storage subsystems directly. On a storage subsystem the configured storage device is assigned directly to a client partition instead of a VIO server partition. The biggest advantage of this method is a simplified assignment of storage devices to client partitions. A client partition will immediately see all disks that are assigned to its virtual fiber channel adapters—no additional configuration steps are required in the VIO server.

One disadvantage can be that the client partitions' virtual fiber channel adapters now have to participate in the SAN zoning again. On the other hand, this brings back the more traditional zoning model that most storage and SAN administrators are familiar with.

Client partitions with NPIV use traditional multipath drivers again. These drivers typically enable load balancing over multiple paths, spreading the load over multiple fiber channel adapters in single or dual VIO server configurations.

Refer to the NPIV section in the IBM PowerVM Virtualization Managing and Monitoring Redbook for detailed information about the design, configuration, and usage of virtual fiber channel adapters on Power Systems:

- ▶ IBM PowerVM Virtualization Introduction and Configuration  
<http://www.redbooks.ibm.com/abstracts/sg247940.html?Open>
- ▶ IBM PowerVM Virtualization Managing and Monitoring  
<http://www.redbooks.ibm.com/abstracts/sg247590.html?Open>

### 5.3.10 VIOS backup

After the VIOS is set up, as well as after any configuration change, you should take a backup of the VIOS.

The `viosbr` command can be used to back up the virtual and logical configuration, listing the configuration, and restoring the configuration of the Virtual I/O Server.

For complete documentation of this command, see:

<http://publib.boulder.ibm.com/infocenter/powersys/v3r1m5/index.jsp?topic=/iphcg/viosbr.htm>

## 5.4 VIO Server monitoring

The VIO Server is integrated into the SAP CCMS monitoring, as far as the operating system is concerned. In this section, we cover the metrics that are available also for monitoring the various I/O components that are available for further integration. This section is based on a simple load example that demonstrated how some of these metrics are captured and



interpreted. This is not an SAP example because SAP systems normally have a far more sophisticated I/O layout.

In this section, we show a subset of the nmon values, which can be useful when you monitor a landscape using Virtual I/O (VIO). The nmon version 12e onwards supports the VIO Server. This overview is based on a functional test that creates non-SAP load on a simple landscape to capture and demonstrate the metrics that are available from nmon.

### Test scenario and setup

On this POWER6-based machine, as shown in Figure 5-12, we have:

- ▶ VIO Server = VIOS 1.5.2.1 that manages shared Ethernet and shared SAN infrastructure.
- ▶ VIO Client = LPAR5 running AIX 6.1 with virtual Ethernet through the VIOS and virtual Disk, which is a SAN LUN on the VIOS.
- ▶ Client has no local disk or direct network connection.

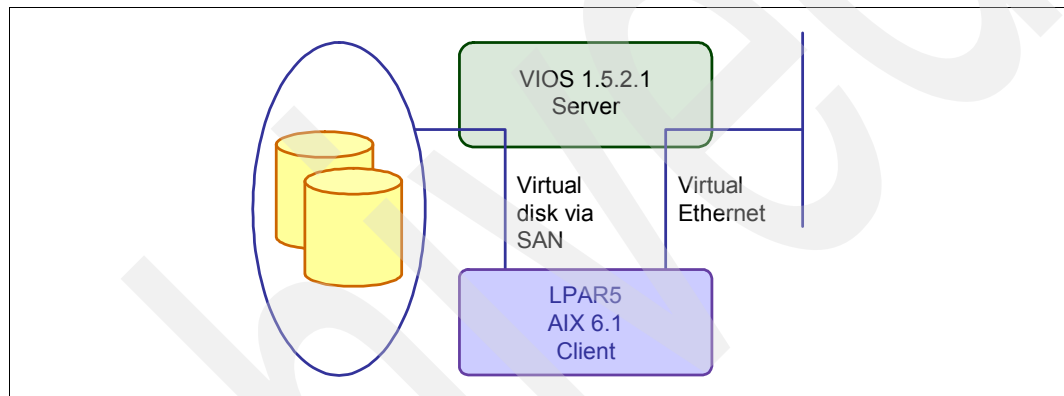


Figure 5-12 Test scenario and setup

The tests were performed in the following manner, as shown in Figure 5-13:

- ▶ Inbound: External machine uses FTP to LPAR5 through the VIO Server, sending a 200 MB file over a physical network to VIO Server and then over the virtual network to LPAR5, where it is written to the virtual disk that is attached through the VIO Server.
- ▶ Outbound: Before this test, the file system that is used in test 1 on LPAR5 is unmounted and remounted to clear the file system cache so that access to this file is from the disk rather than cache. A reverse FTP then caused the file to be read by LPAR5 from disk using the VIO Server and SAN, and then transferred to the external network through the virtual network and VIO Server. FTP was running at about 8 MByte/s, which was limited by the speed of the physical network (100 Mbit/s roughly means 10MByte/s in a maximum).
- ▶ Four programs were started on LPAR5. Each program writes as fast as possible to the LPAR5 virtual disk using the VIO Server and SAN. This is the famous “yes >/home/filename” benchmark.

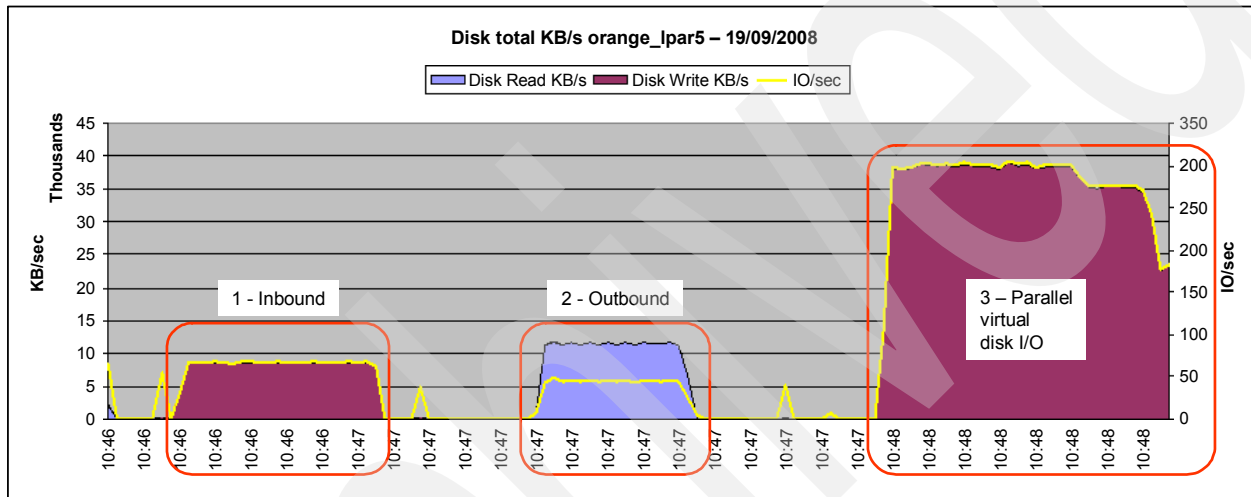


Figure 5-13 View of activities on LPAR5: VIOS client

Referencing Figure 5-13 on page 56 and Figure 5-14:

1. FTP inbound: Data is written to the disk (attached through the VIO Server) but is still visible here as disk I/O. AIX thinks it has a real disk. Only the disk device driver knows differently.
2. Outbound: The data is read from the disk (attached through the VIO Server). The difference in the disk I/O, in the case of reading, is that we can spot the potential of read ahead and AIX changes from 128 KB block in the above test to 256 KB reads. This reduces the number of disk I/O transfers, increases the disk throughput, and reduces the processor required to drive the disk I/O, which is a double win.
3. Four programs writing to disk in parallel (disk attached through VIOS). Instead of a single threaded application, such as file transfer protocol (FTP), after the application is multiple threaded, we can see the opportunity for much higher disk I/O.

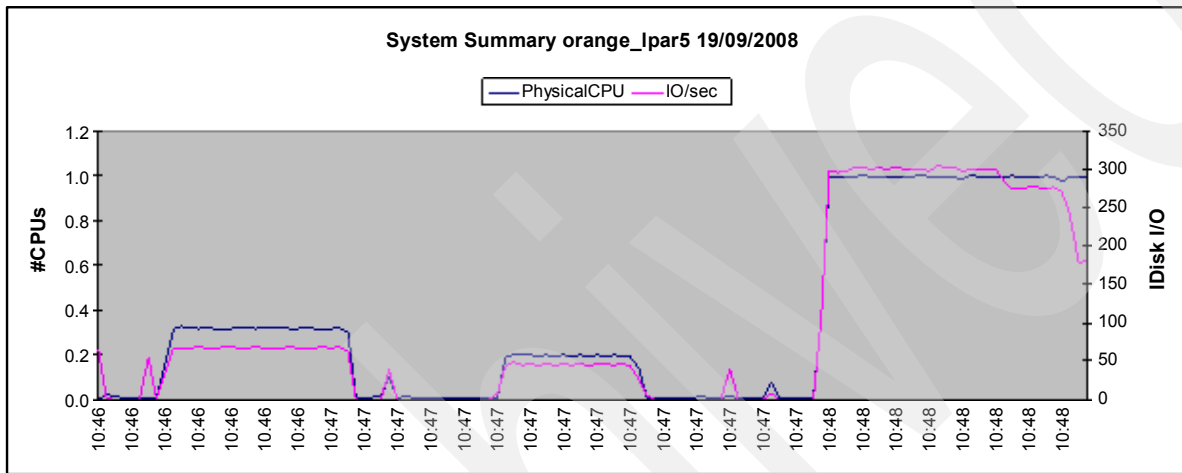


Figure 5-14 Processor usage and disk I/O

Figure 5-15 is a view of the client that shows that the I/O is directly related to the processor time in LPAR5. In this case, there is little application time (“FTP” and “yes” do practically no processing on the data). This highlights that processor time is required to perform I/O operations in terms of AIX kernel and device driver work.

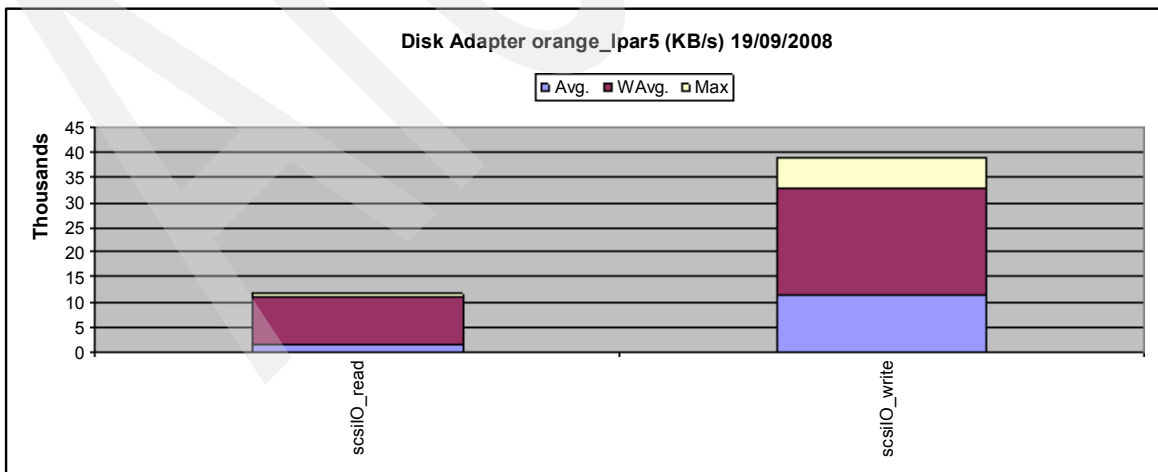


Figure 5-15 Read and write load on disk adapter in KB/s

Figure 5-15 on page 57 shows the disk activities on the virtual SCSI adapter. The blue Average is calculated over the complete time (including the time when no disk I/O occurred). The Weighted Average (WAvg.) only includes the time when the disk was actually busy, which is the number we look at, or the average includes idle time and is artificially low.

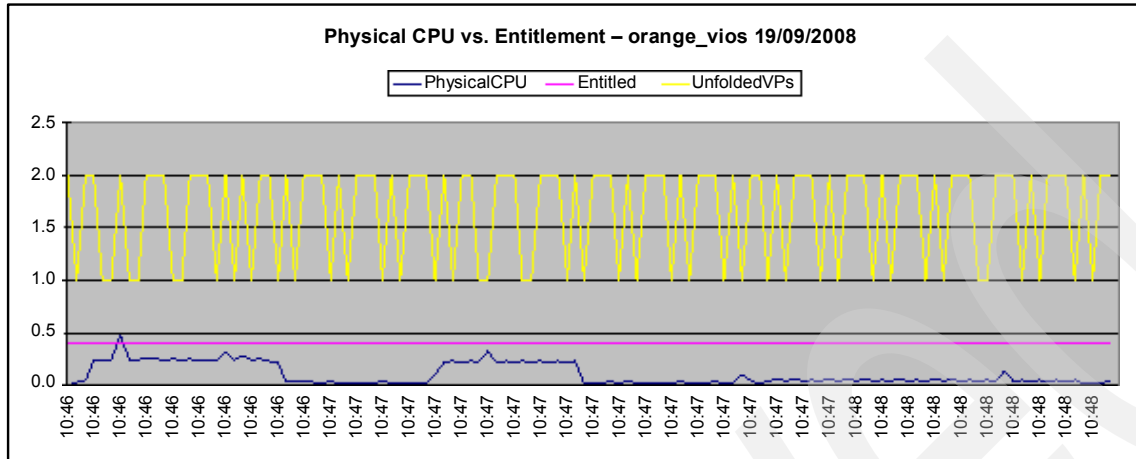


Figure 5-16 View of activities on VIO server

Figure 5-16 shows the number of physical processors on the VIO Server versus the time. For Shared Processor LPARs (SPLPAR) that are uncapped, this is the best number to monitor. The utilization numbers are unsafe because it gets near 100% busy as you get to your entitlement, and you cannot detect when you are using more processor time. Ignore the Unfolded VP line here because it just shows AIX optimizing the number of processors that it is scheduling for maximum efficiency.

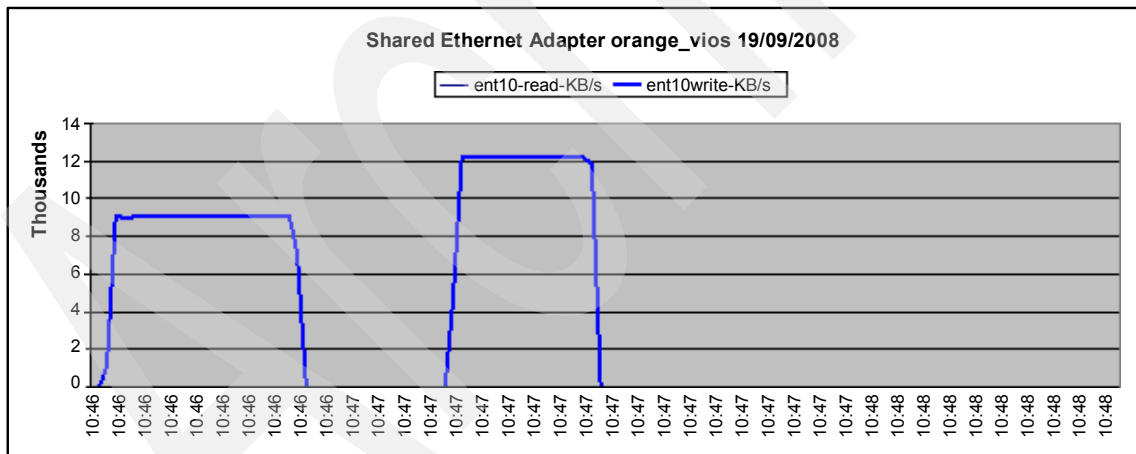


Figure 5-17 Virtual Ethernet I/O read and write throughput

Figure 5-17 on page 58 shows the Virtual Ethernet I/O throughput. The read and write lines are on top of each other, which is because every byte read from a VIOS Client is also written to the physical network. If you are also network connected on the VIO Server (perhaps logged on using telnet or performing an update using FTP) using the SEA (for example, the IP address is not the SEA) then numbers can be slightly different. It is not possible to determine which client is active to what share of the capacity, but we can see what capacity is consumed. Look at the performance data from each VIO client to determine which is doing the I/O. In this case, we only have one active LPAR, so we know who to “blame”.

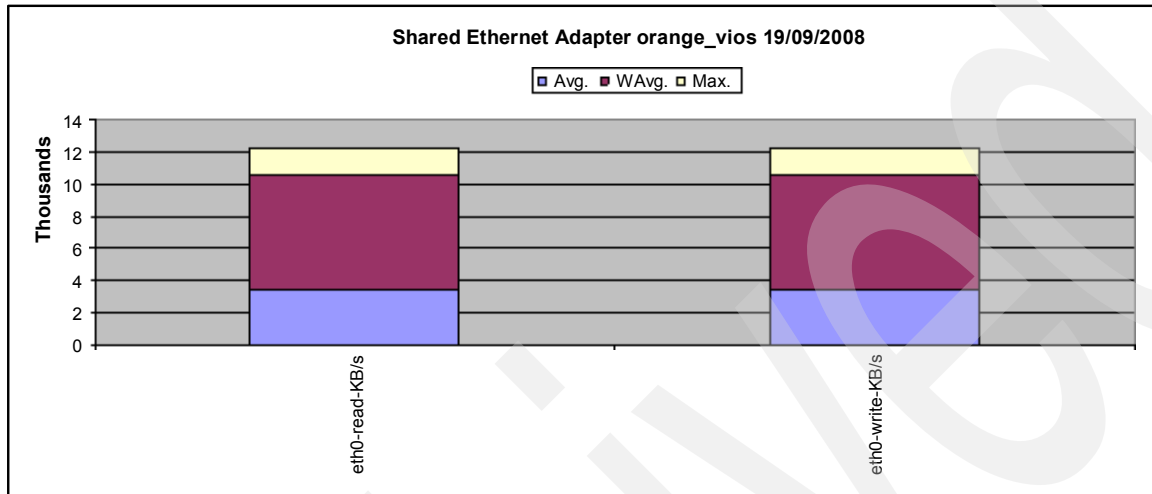


Figure 5-18 Shared Ethernet adapter read and write activities

Figure 5-18 shows the average (Avg.), maximum, and the Weighted Average (WAv.) for both read and write activities over the measurement period for the Shared Ethernet Adapter. When busy, it was just over 10 MBps for both receive and transfer.

Figure 5-19 shows the SEA ent10 and the loopback network pseudo adapter lo0.

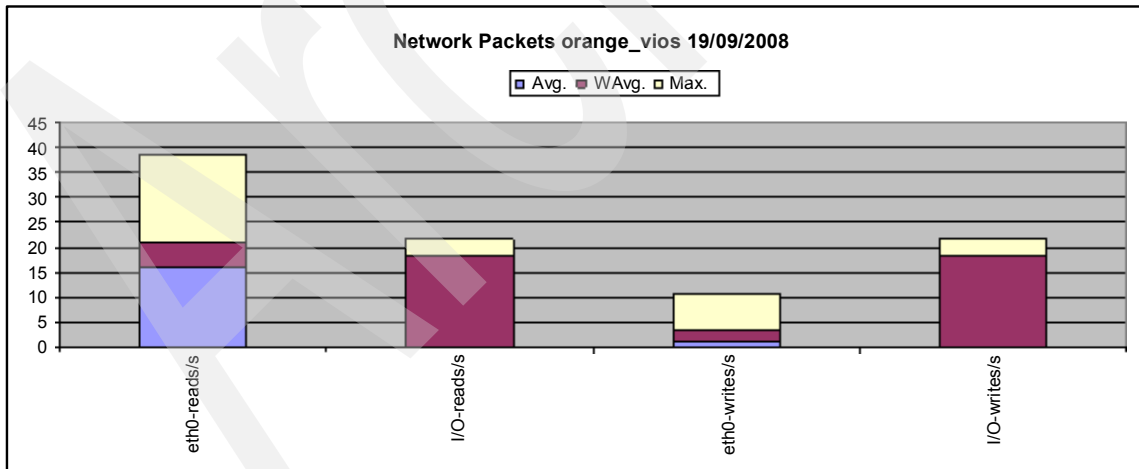


Figure 5-19 Network packets on the Shared Network Adapter and loopback

The number of write packets across the Shared Ethernet Adapter is much less than the number of read packets. Although the same amount of data is transferred, it looks like the VIO Server optimized the network and merged packets. These larger network packets mean that the network is used more efficiently.

Figure 5-20 shows the load on the fiber channel adapter (called fcs0) that is used for VIO Server client LPARs and the internally used SAS adapters (called sissas0) both for read and write.

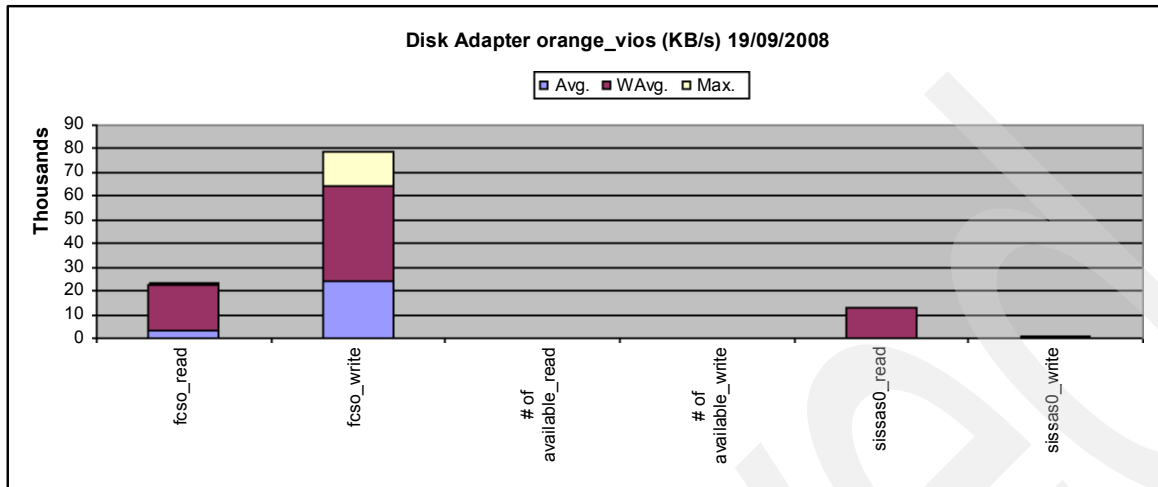


Figure 5-20 Load on fiber channel adapter

The third test was 100% write and thus hit higher disk I/O throughput, which we can see for Weighted Average in the fcs0\_write column on the throughput graph in kbytes per second, as shown in Figure 5-20 and Figure 5-21.

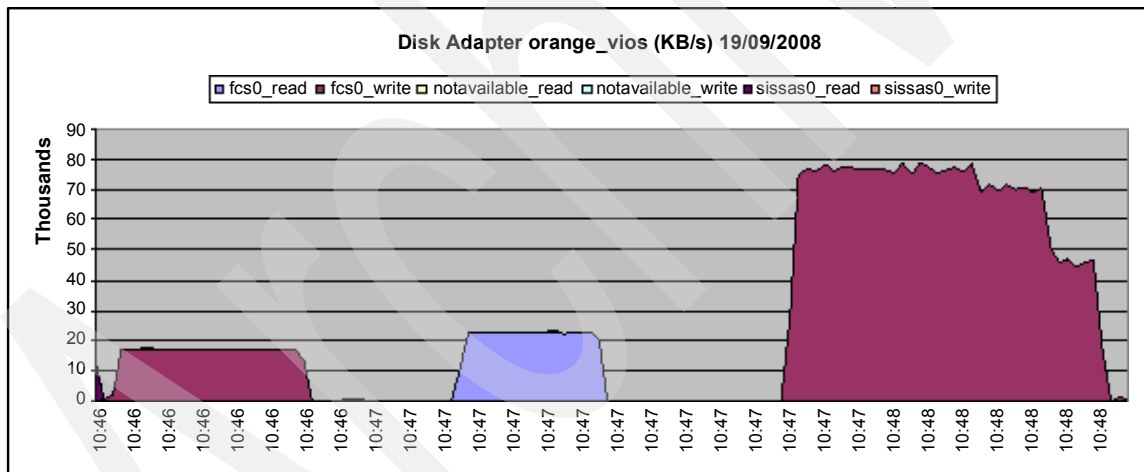


Figure 5-21 Throughput in KB/s on the fiber channel adapter

Figure 5-22 shows an overview of disk I/O on a disk controller. In this simple benchmark case, we only have the one LUN, so the graphs are a lot simpler than in a production type machine.

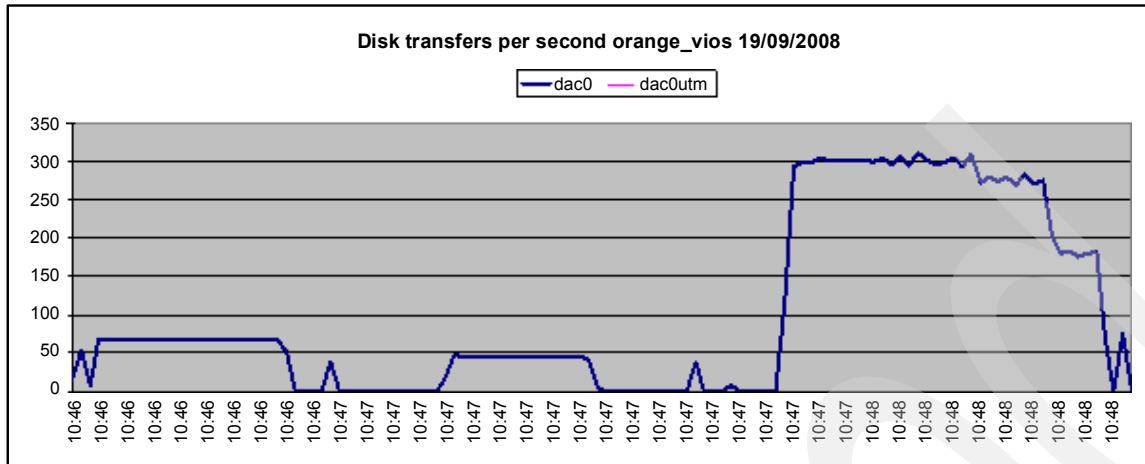


Figure 5-22 Disk transfers per second on the disk controller

Figure 5-23 includes the Service time, which is extremely healthy. Most of the time for this value is between two and three milliseconds.

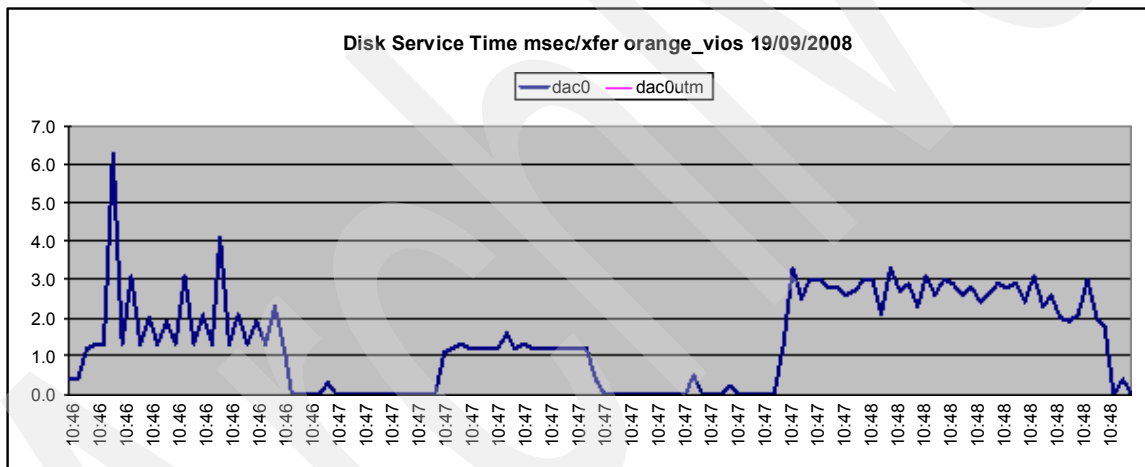


Figure 5-23 Disk service time

Figure 5-24 shows the VIO Server view of disk I/O for each disk. The VIO Server is performing a little disk I/O to the internal disks, but the bulk of the disk I/O is to hdisk6, and the VIO Server client's LUN is running at near 250 transfers per second.

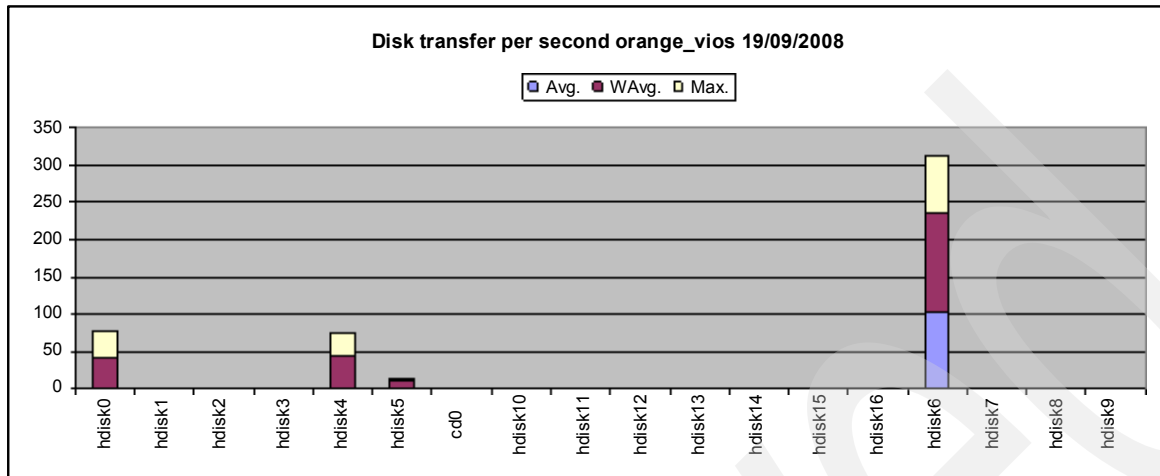


Figure 5-24 VIO server view of disk I/O for each disk





# IBM PowerVM Live Partition Mobility

In this chapter we provide SAP-relevant specifics for the use of Live Partition Mobility (LPM). The International SAP IBM Competence Center (ISICC) provides the technical whitepaper *Live Partition Migration of SAP Systems under Load*, available at:

<http://www.ibm.com/support/techdocs/atsmastr.nsf/WebIndex/WP101917>.

Some parts of this technical brief are also reflected here.

## 6.1 Introduction

Using IBM PowerVM Live Partition Mobility (LPM), SAP clients can move SAP application server and database instances running in AIX or Linux from one physical server to another without the need to schedule application downtime. LPM is currently available for POWER6 and POWER7 technology-based servers.

Even with today's mid-range servers, clients typically run a number of different SAP systems in shared processor pool partitions on a single server. This makes it increasingly difficult to schedule the required downtime window to perform system maintenance or upgrades, as one has to negotiate with a number of different departments. Live Partition Mobility helps to solve this challenge.

Live Partition Mobility can also be used as a mechanism for server consolidation because it provides an easy path to move applications from individual, stand-alone servers to consolidation servers. For 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, one can use LPM 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 and cool machines during off-peak periods. Live Partition Mobility can be automated and incorporated into system management tools and scripts. Support for multiple concurrent migrations makes it possible to vacate systems very

quickly. For single-partition migrations, the HMC offers an easy-to-use migration wizard. LPM can help to achieve the overall continuous availability goals.

With Live Partition Mobility you can eliminate the need to shut down SAP applications to achieve the following goals:

- ▶ Rebalance the total system landscape to adapt to changes in capacity requirements for specific SAP systems.
- ▶ Move SAP workloads from heavily loaded servers to servers that have spare capacity to respond to changing workloads and business requirements.
- ▶ Dynamically move applications from one server to another to avoid planned downtime for:
  - Disruptive firmware upgrades
  - Hardware maintenance or upgrades
  - Moving to new servers during a hardware refresh, for example moving partitions from POWER6 to POWER7 technology-based servers
- ▶ Consolidate workloads and power off unused servers to reduce energy consumption.

A migration transfers the complete system environment. This includes the memory content, processor states, attached devices, and network sessions connected to users or other systems. During the migration process there is a very short period of time where processing is suspended on the source server. LPM then performs some critical synchronization before resuming the partition on the destination server again. To an SAP user this might look like a temporary increase in response time (you would see the hour glass), but running transactions will continue seamlessly after the switch.

Figure 6-1 on page 65 shows a potential migration scenario. A partition called siccpm01 is migrated from an 8-way server to a 32-way server. During the whole duration of the migration process, SAP users that are connected to the SAP application can continue to work.

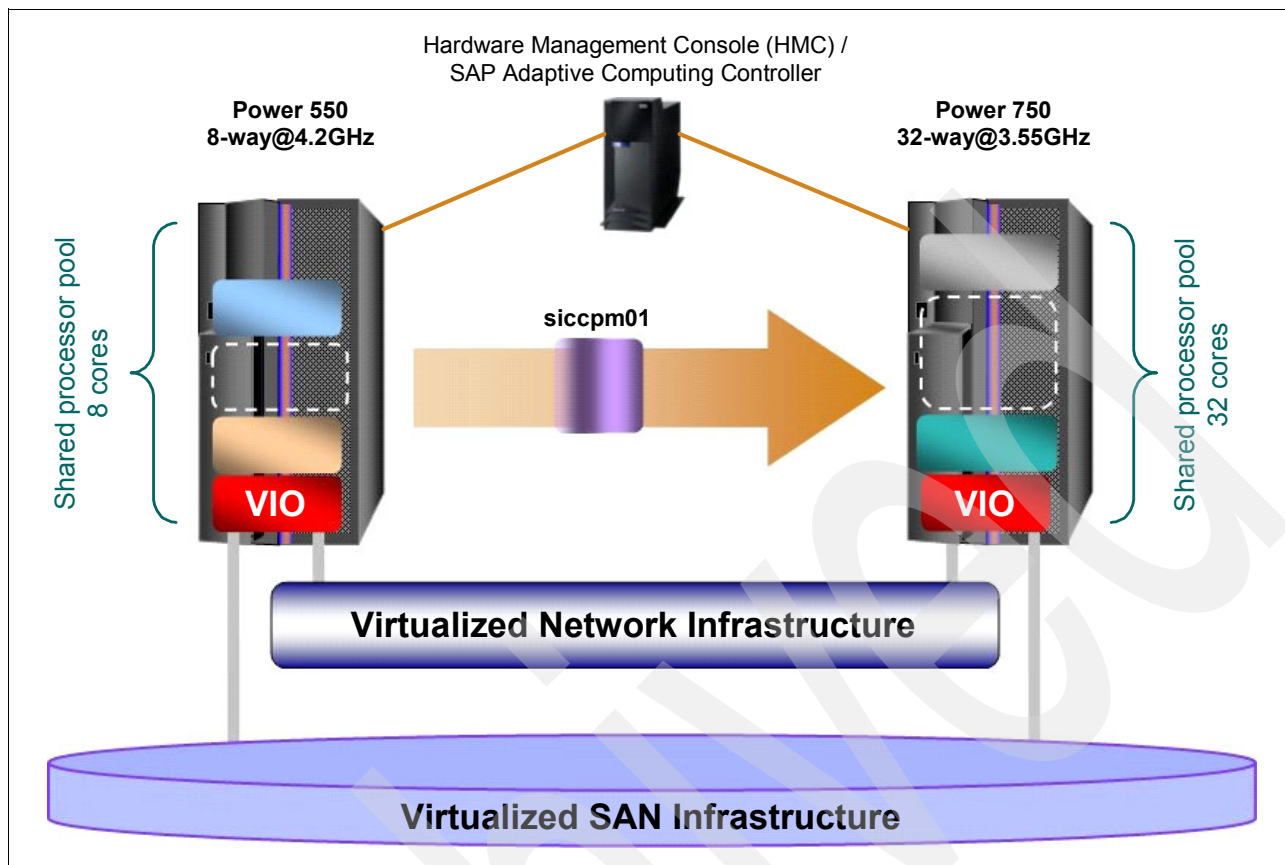


Figure 6-1 Schematic demo setup for LPM in the IBM SAP International Competence Center (ISICC)

It is important to note that Live Partition Mobility is not an alternative for high availability solutions for a number of reasons, for example:

- ▶ Live Partition Mobility requires manual intervention.
- ▶ The HMC governing the migration might not have access to a malfunctioning LPAR and thus might not be able to launch the transfer.
- ▶ The virtualization connections of the VIOs are only stored in the VIO LPARs, but not in the HMC, and those settings cannot be transferred if the server is not available.

## 6.2 Prerequisites

Live Partition Mobility is currently available on POWER6 and POWER7 processor-based servers. AIX 5.3, AIX 6.1, AIX 7.1, and Linux are supported. On IBM i, Live Partition Mobility is planned to be made available with a Technology Refresh for Release 7.1 and will be available on POWER7 processor-based servers. All I/O must be fully virtualized.

The OS and all other data of the mobile partition must reside on external storage subsystems. There must be at least one Virtual I/O Server (VIOS) active on the system and the mobile partition cannot use any physical adapter during the migration. Typical client implementations use a dual VIOS configuration for redundancy. LPM will keep the same configuration when moving a partition to another server (presuming that the target server is also running with a dual VIOS setup).

Requirements, setup, and operations including advanced migration scenarios, as for example the migration of partitions with dedicated I/O adapters, are covered in detail in the product documentation in *IBM PowerVM Live Partition Mobility*, SG24-7460, accessible at:

<http://www.redbooks.ibm.com/abstracts/sg247460.html?Open>

## 6.3 Performance

The ISICC evaluated the performance of the Live Partition Mobility migrations with SAP systems running at various load levels. The results are documented in the whitepaper *Live Partition Migration of SAP Systems under Load*, available at:

<http://www.ibm.com/support/techdocs/atsmastr.nsf/WebIndex/WP101917>

One result documented in that paper is the end-to-end migration rate with different load scenarios (see Figure 6-2). An idle partition can be migrated with a rate of about 4 GB/min. The migration rate decreased to about 2.2 GB/min when the partition was running at about 80% utilization. The transfer rates documented in the paper are consistent with transfer rates clients have achieved with real live migrations of their production systems.

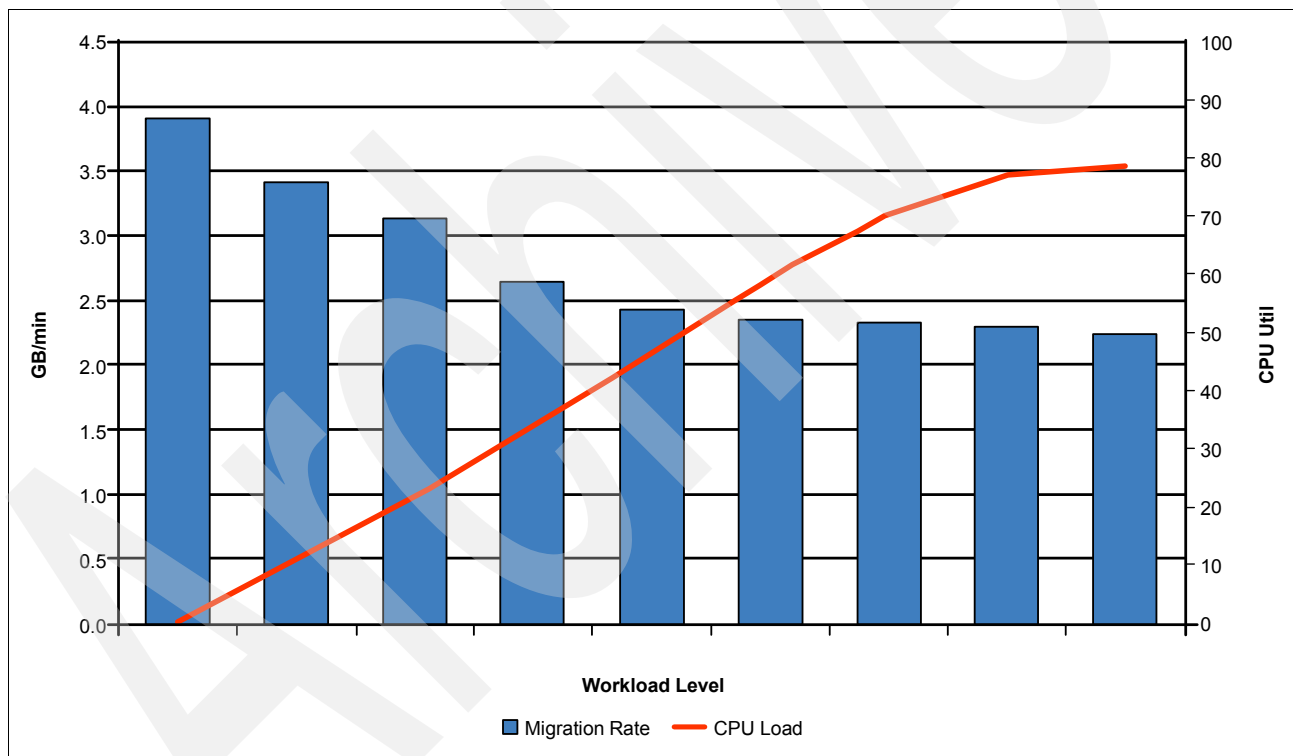


Figure 6-2 Migration rates

## 6.4 SAP notes

SAP-specific recommendations are documented in SAP Note 1102760.

Live Partition Mobility is transparent to SAP applications, and no code changes (for example, in the SAP kernel) are required to exploit this feature. Check the above note for the minimum database levels that are supported with LPM.

## 6.5 SAP license keys

An SAP-specific issue is that SAP license keys (for ABAP and Java systems) are based on a hardware key of the partition that is currently hosting the SAP message server. This hardware key changes after moving an SAP application server instance with a message server to another physical system. The hardware key is only read during start-up of the SAP instance and then buffered in local memory. Because the message server stays up and running during the migration, any subsequent logins will be checked against the old hardware key and will proceed without a problem.

However, the next time the message server is restarted the new hardware key is read and user logins may fail, unless a valid SAP license based on the new hardware key is installed. To avoid any problems regarding this license check, it is recommended to install the additional license keys for all potential target servers in the system landscape in advance. This is similar to what one would have to do for high-availability cluster setups. SAP allows requesting license keys for different hardware keys (check SAP Notes 181543 and 538081).

## 6.6 Flexible License mechanism

To simplify the administration of SAP license keys in system landscapes where the message server can move between several physical servers, SAP has introduced a new license key feature called Flexible License Mechanism. With that method the license key is no longer tied to the hardware key of the message server. The flexible license mechanism uses a separate ID generator, which creates a unique network ID for each message server. This network ID is hardware independent and therefore it is possible to move the message server to a different host and retain the unique ID. It is possible to configure multiple ID generators to eliminate any potential single point of failure.

Refer to the current SAP NetWeaver documentation for a detailed description and instructions of how to set up this new Flexible License Mechanism:

SAP NetWeaver 7.0 EHP2:

[http://help.sap.com/saphelp\\_nw70ehp2/helpdata/en/20/384ef6bd054d5aae1ddd6f67e07a9e/content.htm](http://help.sap.com/saphelp_nw70ehp2/helpdata/en/20/384ef6bd054d5aae1ddd6f67e07a9e/content.htm)

SAP NetWeaver 7.3:

[http://help.sap.com/saphelp\\_nw73/helpdata/en/20/384ef6bd054d5aae1ddd6f67e07a9e/content.htm](http://help.sap.com/saphelp_nw73/helpdata/en/20/384ef6bd054d5aae1ddd6f67e07a9e/content.htm)

## 6.7 Demo of Live Partition Mobility with running SAP systems

There are two videos available that demonstrate the live migration of running SAP systems. The first demo shows the migration of an SAP system running under load using the Hardware Management Console (HMC) to manage the move. The second video shows the integration of IBM PowerVM technology in the SAP NetWeaver stack, making it possible to start LPM operations from the SAP Adaptive Computing Controller (ACC). It also highlights some of the PowerVM-specific performance monitoring metrics that have been integrated in the operating system monitor in SAP CCMS (OS07n).

POWER6 Live Partition Mobility Demo with SAP:

<http://www.ibm.com/support/techdocs/atsmastr.nsf/WebIndex/PRS2921>

IBM PowerVM and SAP ACC Integration demo:

<http://www.ibm.com/support/techdocs/atsmastr.nsf/WebIndex/PRS4232>

## 6.8 POWER6 to POWER7 migrations

Live Partition Mobility can be used to perform hardware refreshes without stopping the application. For example, it is possible to migrate from existing POWER6 processor-based servers to new POWER7 technology-based systems. During such a migration operation, the migrating partition will stay in a POWER6 compatibility mode. This is a partition attribute that can be set in the partition profile. The effect is that after the migration POWER7-specific processor enhancements are disabled; most notably the partition is still running in 2-way SMT (Simultaneous Multi-Threading) mode instead of the new 4-way SMT mode that is available with POWER7 technology-based processors. To fully exploit all of the new POWER7 features, a partition eventually has to be stopped and started in POWER7 mode, which requires a scheduled downtime window.

Note that you cannot perform a live migration of a partition that is running in POWER7 mode back to a POWER6 system. Should this be necessary, you would have to either restart the partition in POWER6 compatibility mode or stop the partition and perform an inactive migration.

## Workload partitions

With the release of AIX 6.1, IBM introduced a new virtualization capability called Workload Partition (WPAR). A WPAR is a software-created, virtualized operating system environment within a single AIX image. Each workload partition is a secure and isolated environment for the application that it hosts. The application in a WPAR “thinks” that it is being executed in its own dedicated AIX instance.

WPARs can be created in all hardware environments that support AIX 6.1 and later. This includes, for example, POWER4 machines, which are supported by AIX 6.1. Figure 7-1 shows that you can create workload partitions within multiple AIX instances on the same physical server, whether they execute in dedicated LPARs or micropartitions.

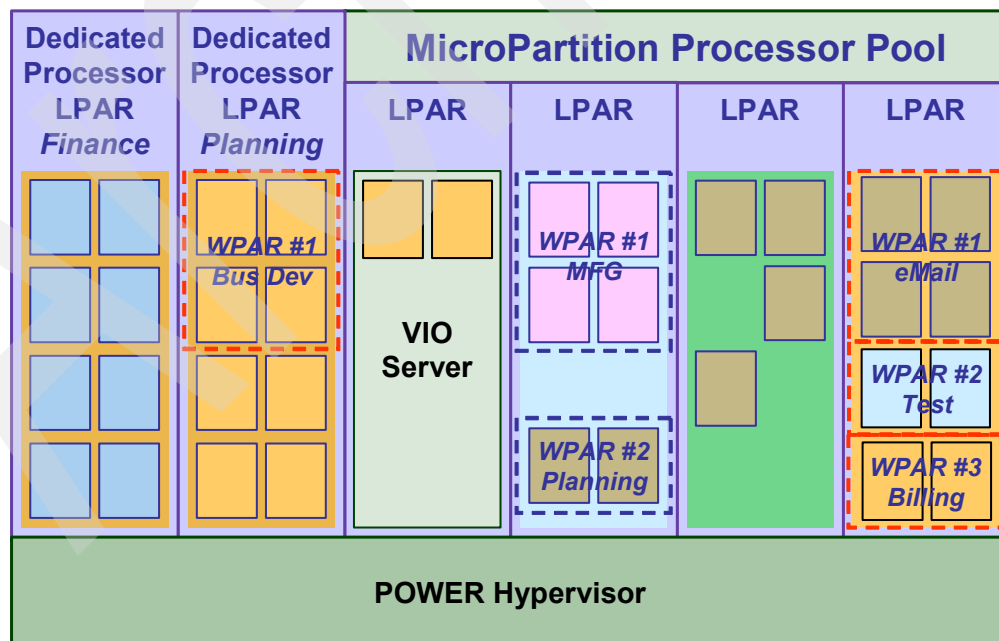


Figure 7-1 WPAR instantiated within dedicated partitions and micropartitions

## 7.1 Characteristics of WPARs

WPARs have the following characteristics:

- ▶ The workload partition technology can help in an environment where an application environment needs to be started often, on-demand, and quickly, which might apply, for example, in test environments.
- ▶ You can use the configuration that is stored in specification files as input to WPAR creation commands, which allows the system administrator to automate, through scripts and programs, the startup and handling of multiple workload partitions.
- ▶ The WPAR technology gives you additional flexibility in system capacity planning as part of a strategy for maximizing system utilization and provisioning efficiency.
- ▶ AIX 6.1 and later provides highly granulated control of processor and memory resource allocation to workload partitions (down to 0.01% increments). This technology is therefore suitable for server consolidation of very small workloads.
- ▶ The WPAR technology allows you to share an AIX instance between multiple applications, while still running each application within its own environment, which provides isolation between applications. In this case, the more applications that are consolidated within one AIX instance, the less the system administrator has to perform operating system (OS) fixes, backups, migration, and other OS maintenance tasks.

## 7.2 Types of WPARs

There are multiple types of WPARs:

- ▶ Application WPARs provide isolation for individual services or applications.
- ▶ System WPARs provide an entire virtualized operating system environment.

Both types can run within a single AIX OS image, which is referred to as the global environment. Starting with AIX 7.1 and POWER7, versioned WPARs are available. Versioned WPARs provide a different OS version in the WPAR than the OS version of the global environment, for example an AIX 5.2 WPAR running in an AIX 7.1.

### 7.2.1 System WPARs

A System WPAR presents a secure and isolated environment that is most similar to a standalone AIX system. Each System WPAR has dedicated writable file systems, although it can share the global environment /usr and /opt file systems in read only mode. Here, we term this difference as Unshared and Shared, respectively.

System WPARs are autonomous virtual system environments and appear, to applications that are running within the WPAR, as though they run in their own separate instance of AIX. Multiple System WPARs can run within the single global AIX image, and each WPAR is isolated from other WPARs.

System WPARs have the following attributes:

- ▶ All typical system services are activated.
- ▶ Operating system services, such as telnet, are supported. You can telnet into a system WPAR as root.



- ▶ Use distinct writeable file systems that are not shared with any other WPAR or the global system.
- ▶ Own private file systems.
- ▶ Own set of users, groups, and network resources.
- ▶ WPAR root user has no authority in the global environment.

The combination of isolation and resource sharing makes System WPARs an excellent feature for consolidating several SAP systems, independent from the underlying POWER hardware. Using a single global AIX image simplifies OS maintenance for multiple SAP systems; however, using one global OS image also introduces a very obvious single-point-of-failure and introduces new dependencies.

Hence, consolidating non-production and smaller SAP instances is expected to be the major use case for this feature.

## 7.2.2 Application WPARs

An Application WPAR has all of the process isolation that the System WPAR does. However, it shares the file system name space with the global system and any other Application WPAR that is defined within the system.

An Application WPAR is essentially a wrapper around a running application or process for the purposes of isolation and mobility. It lacks some of the system services that the System WPAR provides, for example, it is not possible to log in or to telnet into an Application WPAR. When the application that is running in an Application WPAR terminates, the WPAR also ceases to exist.

An SAP system implementation is a complex environment with multiple components, specific requirements to the AIX environment, and many processes. Therefore, Application WPARs are not supported for SAP applications.

For more information about WPARs, refer to *Introduction to Workload Partition Management in IBM AIX Version 6.1*, SG24-7431, available at:

<http://www.redbooks.ibm.com/redbooks/pdfs/sg247431.pdf>

## 7.2.3 Versioned WPARs

Versioned WPARs are intended to provide runtime environments for applications, where the manufacturer of the software does not exist anymore or is not able to provide versions of his software for newer AIX releases. In these cases a client might not be able to update their hardware to a new generation, since the new hardware might not support the OS version required by the old software. Since SAP is providing its software stack very quickly on new versions of AIX and also recertifies older software versions to new AIX versions, SAP does not support versioned WPARs.

## 7.3 Support of SAP applications running in WPARs

You can use all SAP applications, which are based on SAP NetWeaver Application Server for ABAP and JAVA and are available on AIX 6.1 and later, in WPARs. For details, review the SAP Product Availability Matrix at:

<http://service.sap.com/pam>

Detailed information about the current support status can be found in SAP Note 1105456. This note also documents all prerequisites and restrictions and is updated regularly. At the time this book was written the following restrictions and prerequisites applied:

- ▶ The minimum required SAP patch levels are the same as those for the used OS version (for example, see SAP Note 1087498 for AIX 6.1 and SAP Note 1458918 for AIX 7.1).
- ▶ SAP NetWeaver MDM is currently not supported.
- ▶ Supported DB releases (client and server):
  - AIX 6.1 - System WPAR only
    - DB2® LUW v9.1 FP5 and higher
    - DB2 LUW v9.5 and higher
    - MaxDB 7.6 and higher
    - Oracle 10.2.0.4 (no RAC) and higherAdditional requirements:  
AIX 6.1 TL2 SP2 or higher  
APAR IZ52319  
APAR IZ54871  
Set environment variable for DB Instance User:  
**export AIXPERFSTAT\_SCOPE=M** -or-  
**setenv AIXPERFSTAT\_SCOPE M** (depending on the shell you are using)  
Dependencies for the installation see SAP Note 1380038
  - Oracle 11.2.0.1 (no RAC) and higher
- AIX 7.1 - System WPAR only
  - DB2 LUW v9.5 FP6a and higher
  - DB2 LUW v9.7 FP3 SAP and higher
  - MaxDB /liveCache - see SAP Note 720261

### 7.3.1 Supported scenarios

In this section, we provide the scenarios to which the setup is restricted.

#### Unshared System WPARs without resource control

- ▶ New SAP installations:

You can install and run SAP applications in an Unshared System WPAR without any changes required in the SAP installer. Just follow the instructions in the SAP Installation Guide.

- ▶ SAP release upgrade and manual migration scenarios:

SAP applications installed in an Unshared System WPAR can be upgraded with standard SAP upgrade procedures without any restrictions. Just follow the instructions in the SAP Upgrade Guide.

Manual migration scenarios:

The manual migration of an SAP installation from an LPAR into an Unshared System WPAR is supported using the standard SAP Homogeneous System Copy procedure (see the standard SAP System Copy documentation).

#### Shared System WPARs without resource control

- ▶ New SAP installations:

You can install and run SAP applications in a Shared System WPAR without any changes required in the SAP installer, after some WPAR-specific preparations. Review 7.4, “Sample installation of SAP NetWeaver 7.0 in a Shared System WPAR” on page 73. Follow the instructions in the SAP Installation Guide.

## SAP Release upgrade and manual migration scenarios

- ▶ At this time, this scenario is not supported. Tests of this scenario are ongoing. Review the SAP Note 1105456 for current support.

## Versioned WPARs

- ▶ Versioned WPARs are not supported.

## 7.4 Sample installation of SAP NetWeaver 7.0 in a Shared System WPAR

You must first create a Shared System WPAR. You can specify, for example, the Volume Group to be used for the WPAR-specific file systems, the host name (DNS entry exists), and to duplicate the network name resolution from the global environment:

```
#mkwpar -r -g datavg -n is3037w
```

After a few (in our test scenario 2-3) minutes, the WPAR is created, and you can start it with the following command:

```
#startwpar is3037w
```

Prepare the system (global environment and WPAR) for the SAP installation:

1. Increase the paging space in the global environment, as described in the SAP installation guide, because the paging space of the global environment is shared and used by every WPAR of this global environment. For multiple SAP instances that are in multiple WPARs in the same global environment, you must increase the paging space size that is required by each WPAR.
2. In a Shared System WPAR, /usr is a shared file system and read-only from the WPAR. Because SAP installs under the /usr/sap mount point by default, some additional steps are required to allow this. In the global environment, create the mount point /usr/sap and a new file system that is writeable from the WPAR, for example:

- a. Create a new file system for the WPAR per command line or use smitty:

```
#crfs -v jfs2 -g datavg -m /wpars/is3037w/usr/sap -u is3037w -a  
logname=INLINE -a size=5G
```

- b. Mount the /wpars/is3037w/usr/sap file system:

```
#mount /wpars/is3037w/usr/sap
```

3. If you want to install several SAP instances and separate the file systems for each instance, you can create individual file systems for these and mount them under /wpars/is3037w/usr/sap, for example, /wpars/is3037w/usr/sap/NW7
4. Create the other SAP- and DB-specific file systems that are described in the SAP installation guide in the global environment, and mount them under the WPAR-specific file systems, for example:

```
#crfs -v jfs2 -g datavg -m /wpars/is3037w/sapmnt/NW7 -u is3037w -a  
logname=INLINE -a size=5G  
#mount /wpars/is3037w/sapmnt/NW7  
#crfs -v jfs2 -g datavg -m /wpars/is3037w/db2 -u is3037w -a logname=INLINE -a  
size=30G  
#mount /wpars/is3037w/db2
```

5. If you do a 3-tier installation (separate DB host), you must export the directory where the SAP profile and executables are stored, for example, /sapmnt/NW7, from the global environment because there is no NFS server available in the WPAR.
6. The cryptographic software that is used for strong encryption in some SAP solutions (for example, SAP Enterprise Portal) requires policy files from your JDK provider to be installed. Install the unrestricted JCE policy files for the IBM 1.4.2 JDK, which you must do in the global environment because the JDK is installed in the shared /usr file system of the global environment. Otherwise the SAP installer tries to install this, and fails.
7. Download the unrestricted JCE policy files and install them in the global environment. Download from:

<https://www6.software.ibm.com/d1/jcesdk/jcesdk-p>

After this preparation, you can run the SAP installation as described in the SAP installation guide without any restrictions.

## 7.5 Live Application Mobility

Live Application Mobility requires an additional Licensed Program Product called Workload Partitions Manager™ for AIX.

The Workload Partitions Manager provides more advanced features for managing multiple WPARs, including:

- ▶ Graphical interface with wizards for many management tasks
- ▶ Resource allocation
- ▶ Role-based views and tasks
- ▶ Centralized, single point-of-administration
- ▶ Enablement for Live Application Mobility
- ▶ Automated application mobility based on policies

To transfer an application to another LPAR or server, a checkpoint/restart feature is used. During a WPAR checkpoint, the current state of running applications is saved and during restart operations resumed in the new AIX image.

With this feature, it is possible to move WPARs between compatible systems without significantly affecting WPAR users.

Live Application Mobility is not supported in SAP environments. Review SAP Note 1105456 for current support.



## SAP system setup for virtualization

In this chapter, we discuss:

- ▶ SAP Adaptive Computing
- ▶ SAP instance profile changes
- ▶ Virtual memory tuning in AIX for SAP systems
- ▶ Main storage pools, work process priorities and workload capping on IBM i
- ▶ Active Memory Expansion (AME)
- ▶ processor utilization metrics

## 8.1 SAP Adaptive Computing

The central idea of SAP Adaptive Computing builds on the virtualization of the following layers: services (SAP application components), computing resources, and a main storage device, as shown in Figure 8-1. This complements the undisputed virtualization technologies of IBM PowerVM, which leads to increased flexibility and easier management of planned downtimes.

SAP covers the subject on the SAP Community Network (SDN) at:

<https://www.sdn.sap.com/irj/sdn/adaptive>

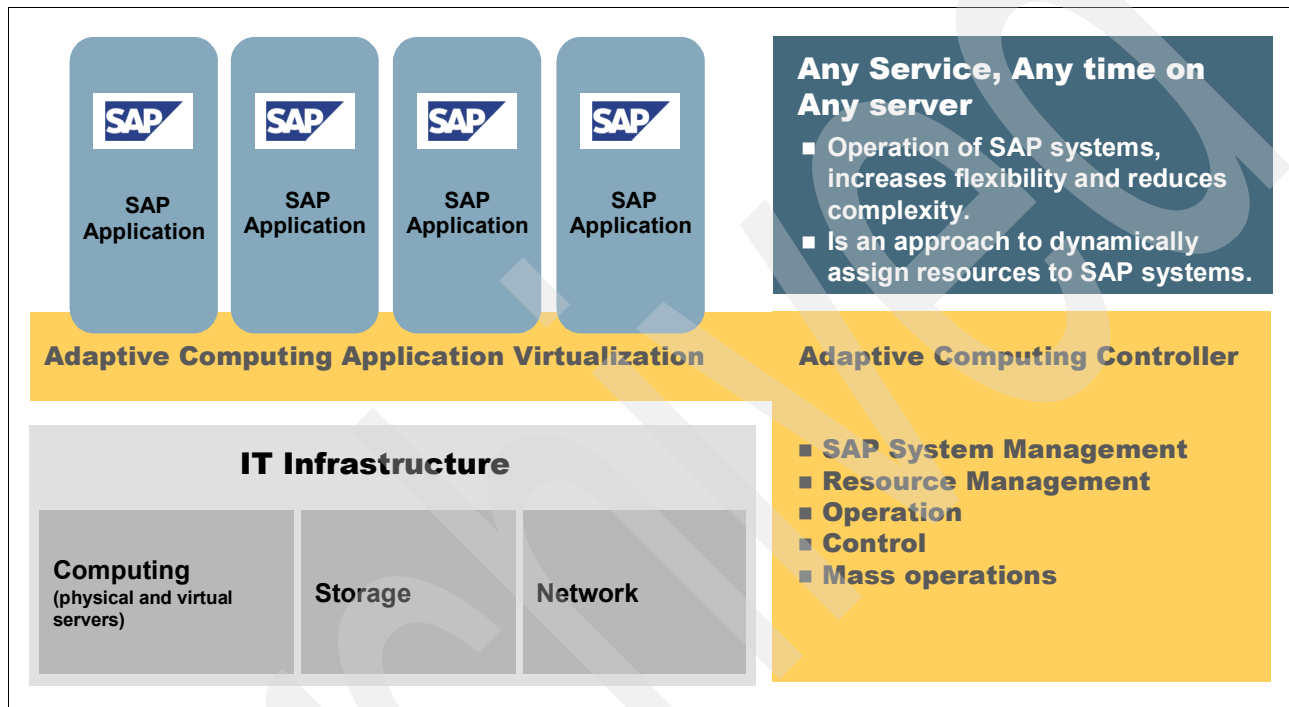


Figure 8-1 SAP Adaptive Computing concept<sup>1</sup>

### 8.1.1 Overview

One challenge in today's IT environments is to avoid an increasing Total Cost of Ownership, which is why SAP developed a concept that has the central objective to "start any service on any server at any time". The Adaptive Computing infrastructure is an integral component of SAP NetWeaver.

A service, in this respect, is an SAP product component, for example, the central system of SAP ERP, an application server in a BW landscape, or a DB server.

The concept consists of the following four building blocks:

- ▶ **Computing (physical and/or virtual):** Standard servers or LPARs here are called compute nodes. Also recommended is a central location to perform operating system maintenance and image distribution. In the case of AIX, this is done with the Network Installation Manager (NIM).

<sup>1</sup> Copyright SAP AG 2010. All rights reserved.

- ▶ **Storage:** The main storage pool is set up as a SAN or NAS concept. The compute nodes should not contain local disks. Only the operating system and paging space can reside on local disks, if desired. At least all of the application-relevant data must reside on a storage device that is external to the server. Recommended storage solutions that substantially simplify the storage administration are distributed file systems (such as, IBM General Parallel File System) or Network Attached Storage solutions.
- ▶ **Network:** High-speed network connection between the servers and the storage (for example, Gigabit Ethernet)
- ▶ **Control:** The Adaptive Computing Controller (ACC) is an infrastructure-management solution that communicates with different components of the infrastructure.

The ACC is a Java-based application that manages the resources (servers and services) that are known in a system landscape. From the ACC an application service (for example, a central instance of an SAP application component) can be started, stopped, or moved from one compute node to another. This is done with disruption with the classic ACC functionality called relocation.

It is also possible to trigger a Live Partition Migration from the SAP ACC. The interface to the IBM HMC or the Systems Director is available and here the move happens without disruption, without loss of service. Details about the interfaces are available in the following SAP Notes:

1411300 - *Configuration of IBM HMC for Power Systems Adapter for ACC*

1539332 - *Configuration of IBM Systems Director VMControl for ACC*

They are located on the SAP Support Portal at:

<http://service.sap.com/support>

The supported (managed) systems in the system landscape can be 4.6C systems (using the 4.6D kernel) and higher. The management of older SAP releases is not supported.

SAP Adaptive Computing is an infrastructure implementation concept, which is realized by the partners of SAP with their technology. SAP provides the SAP ACC as an administration console to the solution.

## 8.1.2 Adaptive Computing compliance test

Adaptive Computing compliance test is a quality assurance procedure to ensure that clients get solutions that are properly tested and work. It is not a certification, hence support is provided by SAP as long as generally supported HW and SW components are used in an adaptive system setup. The building blocks of the compliance test are: the version of the ACC, the storage technology, and the technology of the computing nodes. A list of compliance tests is available on the SAP Community Network (SDN) at:

<https://www.sdn.sap.com/irj/sdn/adaptive>

## 8.1.3 Technical Implementation

A cookbook about how to implement SAP Adaptive Computing with IBM Power Systems is available on IBM TechDocs at:

<http://www.ibm.com/support/techdocs/atsmastr.nsf/WebIndex/WP100914>

## 8.2 SAP instance profile changes

In this section, we cover dynamic work processes and extended memory tuning recommendations.

### 8.2.1 Dynamic work processes

In the past, adding or removing SAP work processes to or from an application server instance required a restart of the instance. In a running instance you were only able to change the type of already existing work processes. For example, you could define operation modes to switch a certain number of batch processes to dialog processes and vice versa. One of the latest features that SAP ships are dynamic work processes. To use dynamic work processes, you must initially create them by changing the instance profile and restarting the application server instance.

After they are created, the system owns dormant work processes that can be activated and deactivated while the system is running by:

- ▶ The system itself to resolve deadlocks
- ▶ The administrator to adapt the system to changing load or to address changing resource availability

This feature is especially valuable for virtualized environments where the amount of processors per LPAR can be dynamically changed. With dynamic work processes, dormant work processes can be predefined and they will not use processor resources while remaining in that state. Now the administrator of the SAP system can easily address changing hardware situations while the SAP system stays up, which was not possible without a restart before. This feature requires at least SAP Kernel 7.10.

To learn how to configure dynamic work processes, go to:

[http://help.sap.com/saphelp\\_nwpi71/helpdata/en/46/c24a5fb8db0e5be1000000a1553f7/frameset.htm](http://help.sap.com/saphelp_nwpi71/helpdata/en/46/c24a5fb8db0e5be1000000a1553f7/frameset.htm)

### 8.2.2 Extended memory tuning recommendations for AIX

PowerVM technology allows the available resources for an LPAR to vary dynamically over time without restarting the partition. As already described, this can occur for processor resources, for example, through properly configured shared processor pool LPARs and for memory with DLPAR operations. Therefore, a partition can adjust the processor capacity to the actual workload on the system.

In this section, we describe the recommended SAP memory settings to accommodate this ability on AIX. The SAP Application Server ABAP stores all user-specific data in a so-called user context. To allow the user to be scheduled to any work process in the system, this context resides in shared memory. On AIX, two implementations are available to use shared memory for SAP systems: SAP memory management allows you to use different implementations by setting instance profile parameters; on AIX, the recommended implementation is sometimes referred to as the SHM\_SEGS implementation. Because this is not the default setting of the SAP kernel, ensure that the following values are set in the instance profile:

```
ES/TABLE = SHM_SEGS  
ES/SHM_SEGS_VERSION = 2 (see also SAP Note 856848)
```



With this setting, each user context is stored in its own shared memory segment. This segment gets attached to a work process only when the user is dispatched to a work process. Hence the user data is isolated. In case the user logs off from the SAP system, its context is not needed any more, and the memory can be released to the operating system. In the current implementation of the SAP memory management algorithm, there are several optimizations, such as caching of released shared memory segments to optimize performance. Over time the number of shared system memory segments in the system adjusts to the number of users that are active in the SAP system, and thus the memory usage adjusts to the current workload of the SAP system.

A summary of the advantages is as follows:

- ▶ Caching of segments by non-OS mechanisms to reuse them instead of returning them directly to the operating system.
- ▶ Improvement of handling large user contexts and the possibility to limit the number of segments per user.
- ▶ Improvement of fast context switches.
- ▶ Deferred page space allocation, which allows you to allocate just enough paging space to serve the amount of currently used extended memory segments instead of allocating paging space for the maximum that you can use.

It is highly recommended that you use the optimized SAP memory management implementation on AIX. Table 8-1 lists the related SAP Notes:

Table 8-1 SAP Notes covering memory management on AIX

SAP Note	Title
856848	AIX Extended Memory Disclaiming
1088458	AIX: Performance improvement for ES/SHM_SEGS_VERSION
191801	AIX 64-bit with very large amount of Extended Memory (refers to SAP notes 445533 for 4.6D and older releases and SAP Note 789477 for 6.10 and newer releases)
1551143	SORT (AIX): Memory release
973227	AIX Virtual Memory Management: Tuning Recommendations
1121904	SAP on AIX: Recommendations for paging space

### 8.2.3 Extended memory tuning recommendations for IBM i

Beginning with SAP kernel release 6.40, there are two possible ways to allocate memory in an SAP Application Server ABAP: *Memory Mapped to File* and *Shared Memory*. Early performance tests had indicated that the method *Memory Mapped to File* was significantly faster than *Shared Memory*, so *Memory Mapped to File* was chosen as the default setting for all SAP systems running with a kernel release 6.40 and higher.

In IBM i 6.1, significant enhancements were made to storage management, so that the method *Shared Memory* now became faster than the method *Memory Mapped to File*. Nevertheless, the method *Memory Mapped to File* remains the default for two reasons:

- ▶ The performance of *Shared Memory* only affects *Roll-In* and *Roll-Out* of user contexts in work processes. Because other transaction components are responsible for a much higher share in the response time, the overall performance improvement of *Shared Memory* is very small.

- ▶ Prerequisites are: IBM i 6.1, SAP kernel release 7.10 or higher with certain patch levels, and the UNIX style operating system user concept as described in SAP Note 1123501.

To switch to method Shared Memory, ensure that all prerequisites are met and set the following profile parameter (all uppercase):

```
ES/TABLE = SHM_SEGS
```

You can find the complete list of prerequisites and additional profile parameters in relation to this setting in SAP Note 808607. This SAP Note references SAP Note 789477 for additional tuning information on systems with very many concurrent users. Even though the title of SAP Note 789477 points to AIX as operating system, you can use the recommendations for IBM i as well.

### 8.3 Virtual Memory tuning in AIX for SAP systems

The memory management of AIX is highly optimized to use the installed physical memory of the server in a very efficient way. Looking for the free memory of a running AIX system typically shows small values for this metric, even without significant load from an application program.

Example 8-1 shows the output of the `vmstat` command. The column `fre` indicates about 4.8 MB of free memory. Small values like this often lead to irritations, assuming that the system is running out of free memory for the running application, such as an SAP system. In the next section, we describe in detail how the memory management of AIX works and why a small value for free memory does not directly indicate a critical system situation. We also give recommendations for the parameter setup of the memory management.

*Example 8-1 Output of the vmstat command*

---

```
#vmstat -w 1 5
System Configuration: 1cpu=2 mem=22528MB
```

kthr		memory		page						faults			cpu			
r	b	avm	fre	re	pi	po	fr	sr	cy	in	sy	cs	us	sy	id	wa
1	0	4490108	4930	0	0	0	0	0	0	16	189	250	0	0	99	0
1	0	4490108	4930	0	0	0	0	0	0	16	104	247	0	0	99	0
1	0	4490107	4931	0	0	0	0	0	0	16	160	271	0	0	99	0
1	0	4490107	4931	0	0	0	0	0	0	15	109	244	0	0	99	0
1	0	4490107	4931	0	0	0	0	0	0	9	189	277	0	0	99	0

---

In this section, see SAP Note 973227, which is located on the SAP Support Portal at:

<http://service.sap.com/support>

AIX Virtual Memory Management (VMM) is designed to exploit all available physical memory in a system. VMM distinguishes between computational memory segments (working storage segments, such as process data and stack or program text segments) and file memory segments (usually pages from permanent data files). If the virtual memory demand can be contained in real memory pages, VMM fills up the available pages with computational memory pages or file pages (as a result of I/O operations). When the number of free pages drops below a certain threshold, a page replacement process starts and attempts to free up memory pages. The page-replacement algorithm uses a number of thresholds and repaging rates to decide which pages get replaced. With the default threshold settings on AIX 5, there is a slight bias in favor of computational pages. Especially for database servers, the default settings are not optimal.

Databases buffer access to persistent data in their own cache, and as the data is typically accessed through file system I/O operations, it is also buffered by the VMM in file pages. This redundant buffering of the same data can lead to unnecessary paging. The objective for SAP systems is to keep all working storage segments in real memory, while using the unused memory segments as file system cache. Page replacement should only steal memory pages from the file system cache.

With older AIX releases, the only way to achieve this objective was to set certain thresholds, such as `minperm`, `maxperm`, and `maxclient`, to rather low numbers (see SAP Note 921196).

These parameters do not set the amount of memory that is used for file system cache; instead, they control the boundaries for the page replacement algorithm. If the page stealer must free up memory and the amount of memory used by file pages is between `minperm` and `maxperm` (or `maxclient` for JFS2 file systems), the page stealer considers the repaging rates to determine which type of pages to steal. If the repaging rate for file pages is higher than the rate for computational memory, the page stealer steals working segments, which leads to the undesired paging for SAP systems.

Later releases introduced a new parameter called `lru_file_repage` that allows you to turn off the check for repaging rates. With `lru_file_repages` set to 0, the page replacement algorithm always steals file pages if the amount of memory that is used for file pages is larger than the `minperm` setting. In AIX 6.1 this parameter is set by default to 0 and in AIX 7.1 this parameter has been removed, but the system is behaving as if the parameter would have the value 0. With this new parameter, the recommendations for VMM page replacement tunables are:

- ▶ `minperm%` = 3 (default 20)
- ▶ `maxperm%` = 90 (default 80)
- ▶ `maxclient%` = 90 (default 80)
- ▶ `lru_file_repage` = 0 (default 1)
- ▶ `strict_maxclient` = 1 (default 1)
- ▶ `strict_maxperm` = 0 (default 0)

You can set the parameters using the smith fastpath “TunVmo” or directly using the following command:

```
/usr/sbin/vmo -p -o minperm%=3 -o maxperm%=90 -o maxclient%=90 \  
-o lru_file_repage=0 -o strict_maxclient=1 -o strict_maxperm=0 \  
-o minfree=960 -o maxfree=1088
```

The new parameter recommendations are not restricted to DB servers only and can be implemented for all servers in the landscape (including application servers in a 3-tier implementation). In AIX 6.1, the recommended settings are the default settings for each new AIX installation.

With this configuration, the system can use up to 90% of its memory for file caching but favors computational pages over file pages. Page replacement does not steal any computational pages unless the amount of computational memory exceeds 97%.

There are two more tunables that control the behavior of the page replacement algorithm: `minfree` and `maxfree`. The VMM attempts to keep the size of the free list greater than or equal to `minfree`. When page faults or system demands cause the free list size to fall below `minfree`, the page-replacement algorithm runs and frees pages until the number of pages on the free list reaches `maxfree`. The default values for `minfree` and `maxfree` were increased in AIX 5.3 ML01 and can be implemented for older releases too. The new default values are:

```
minfree = 960  
maxfree = 1088
```

These settings are per memory pool. Larger systems can have more than one memory pool (you can check the number of memory pools with the `vmstat -v` command). In most cases, the defaults are fine and do not require specific tuning. The exceptions are typically small systems or LPARs with a single memory pool and heavy file system activity. Larger systems have multiple memory pools and the system wide minfree and maxfree values are large enough to prevent the depletion of the list of free memory pages. If you see the free list dropping close to 0 (check the fre column in a `vmstat` output), increase the minfree and maxfree values by 1024 and monitor again.

The recommendations require the following minimum AIX levels for this parameter set to work correctly. If you are on an older level, continue to use the recommendations from SAP Note 921196. The required maintenance levels are:

- ▶ AIX 5.1 ML09
- ▶ AIX 5.2 ML05
- ▶ AIX 5.3 ML01

Check the latest requirements for the maintenance level of the AIX version used.

## 8.4 Main storage pools, work process priorities, and workload capping on IBM i

As we point out in Chapter 1, “From a non-virtualized to a virtualized infrastructure” on page 1 of this book, consolidation is an important practice to reduce complexity in the IT landscape. On IBM i, the easiest way to consolidate is to run multiple SAP systems in a single logical partition or on a single server. The operating system and Licensed Internal Code provide system management functions to assign hardware resources such as processor time and main storage pages to requesting processes automatically.

In general, you do not need to configure much—the operating system algorithms are designed to automate workload dispatching as much as possible. Processor dispatching is mainly controlled through the run priorities of the processes or threads, and through the activity level in a main storage pool. The Storage Management component in the Licensed Internal Code ensures that memory pages are allocated in main storage and on disk, or read from disk as they are needed. To make room for the new pages, pages that have not been used for a while are being purged from main storage. For many cases this concept is working very well, but there are certain SAP landscapes where you should consider a special setup.

### 8.4.1 Separation of main storage pools

The IBM i operating system is running with a minimum of two main storage pools: The \*MACHINE pool for Licensed Internal Code tasks and system processing, and the \*BASE pool for all application processes. In addition, you can configure additional main storage pools for special purposes, either *private pools* to hold certain objects permanently in main storage, or *shared pools* to separate the main storage used by a certain group of processes from the remaining processes. Private pools can be assigned to only one subsystem, while shared pools can be used by multiple subsystems at the same time. When using SAP software on IBM i, it can be helpful to set up shared main storage pools for some SAP instances or processes in the following cases:

- ▶ When combining SAP Application Servers ABAP with SAP Application Servers Java in one IBM i partition, you should separate the ABAP and Java servers into multiple shared main storage pools. The SAP Application Server Java is starting so-called *garbage collections* on a regular basis. During the execution of these garbage collections a lot of

pages are purged from the main storage pool. This may affect the performance of other application servers running in the same main storage pool.

- ▶ When running multiple SAP systems in one IBM i partition and one of them is used rarely, it may be advisable to run that one in a separate main storage pool. If all SAP systems are sharing the same main storage pool, the more active SAP systems will purge many pages of the rarely used system from main storage. When the rarely used system is then used again, all these pages need to be paged in from disk, so the response times on the first transactions can be very long.

SAP Note 1023092 describes how to set up shared main storage pools on IBM i. Although the short text of the SAP Note is saying “Using Separate Memory Pools for *Java* Processes”, the section “Assign a memory pool to an entire SAP instance (Java standalone)” can be used for ABAP instances as well.

You can use SAP and IBM sizing guidelines in order to decide what sizes are needed for the separate main storage pools. You can also find some guidance in SAP Note 49201 for the SAP Application Server Java and in SAP Note 808607 for the SAP Application Server ABAP (SAP release 6.40 and higher).

## 8.4.2 Work process priorities and workload capping

Processes that are currently not in a wait state are queued in the task dispatching queue in order to get access to the available processor resources. They are processed in the order of their run priority (lower run priority value = higher priority) and arrival times in the task dispatching queue. Processes are removed from the processor when they are reaching a wait state (such as a lock wait or a wait on a disk operation), or when they reach the end of their time slice, which is defined in the process class. Traditionally, interactive processes run with higher priority and shorter time slice, while batch processes have lower priority and a longer time slice. In contrast, on a typical SAP installation all work processes of all instances are running at the same priority and time slice, so all processes compete against each other. This can cause problems in the following situations:

- ▶ When using dialog work processes 24 hours per day through all time zones instead of having distinct “day” and “night” operation modes, you cannot separate background work from interactive work by the time of the day. In this case, it can be helpful to assign different run priorities to dialog work processes and background work processes.
- ▶ When running multiple SAP systems in a single IBM i partition and one of the SAP systems is using a significant amount of processor capacity, this can have a negative impact on the remaining SAP systems. For example, when an SAP instance is started, each work process requires a significant amount of processor capacity to set up its runtime environment. When an instance with many work processes is started on a partition with few processors, it can keep all available processors occupied for several seconds or even minutes.

SAP Note 45335 describes how to configure different run priorities for different types of work processes. By default, all SAP processes have a run priority of 20, which is the same as value 'M' for the relevant profile parameter `rdisp/prio/<type>` where `<type>` can be `upd`, `btc`, `spo`, `gwr` or `java`. For `<type> = java` you can select higher or lower priority than the rest of the instance, while for the other types you can only choose lower priority than the other jobs. In addition, there is a profile parameter, `rdisp/prio/wp/start`, to change the run priority just during the startup of the instance. To change the run priority of all work processes of an instance, you can also change the run priority in class `R3<sid>/R3_<nn>`, where `<sid>` is the SAP system ID and `<nn>` is the instance number. This can be useful if an unimportant test system and an important productive system share the same partition. Note that a higher number for the run priority value means a lower run priority and vice versa.

Workload capping is supported in IBM i 6.1 with PTF SI41479 or IBM i 7.1 with PTF SI39795 and higher. It allows to restrict the total processor utilization of jobs or subsystems. Each SAP instance is running in its own subsystem, so you can use workload capping to limit the number of virtual processors that can be used by this instance. The disadvantage of workload capping is that it cannot be monitored in the SAP tools for operating system performance. The SAP operating system performance tools (executable saposcol and transactions ST06, OS06 or OS07) are monitoring the overall processor utilization in a logical partition, but they cannot monitor processor utilization per subsystem. If a logical partition has four physical and virtual processors available and an instance subsystem is limited to two virtual processors by workload capping, the SAP performance tools are only showing a processor utilization of 50 %, even though the SAP instance cannot use more processor resources.

To set up workload capping, first create a workload capping group. This workload capping group defines a maximum number of processors that can be used by the associated subsystems. For example, to create a workload capping group named SAPTEST with a maximum of two processors, use the command:

```
ADDWLCGRP WLCGRP(SAPTEST) PRCLMT(2)
```

The next step is to create a character type data area named QTWLCGRP in library QSYS. To create the data area, enter the following command:

```
CRTDTAARA DTAARA(QSYS/QTWLCGRP) TYPE(*CHAR) LEN(2000)
```

The data area contains a list of subsystem description names with their associated workload capping group names in pairs of 10 character names each. As an example, assume that you have a partition with four virtual processors. In the partition, a productive system is running in instance 00 (subsystem: R3\_00), and a test system is running with instance 50 (subsystem: R3\_50). You want the productive system to use no more than three processors, and the test system to use no more than two processors, so you have created a workload capping group SAPPROD with PRCLMT(3) and a workload capping group SAPTEST with PRCLMT(2). To set up workload capping for this set, modify the contents of the data area with the following command:

```
CHGDTAARA DTAARA(QSYS/QTWLCGRP)
           VALUE('R3_00 SAPPROD R3_50 SAPTEST')
```

**Note:** The names must be filled with blanks up to 10 characters. The changes take effect only after the affected subsystems are ended and started again.

When you start the SAP instances with these settings, the job logs of the subsystem control jobs will contain the message CPI146C “Subsystem <subsystem name> is using workload capping group <workload capping group name>”. Use the DSPWLCGRP command to display the processor limit that was defined for all or a specific workload capping group. More information about workload capping groups can be found in the IBM i and System i® Information Center at:

<http://publib.boulder.ibm.com/eserver/ibmi.html>

Select **i 7.1** and follow the path: **IBM i 7.1 Information Center** → **Systems management** → **Work management** → **Managing work** → **Managing workload capping**.

## 8.5 Active Memory Expansion for SAP systems

The IBM POWER7 technology-based systems with AIX provide the new feature Active Memory Expansion (AME), a technology for expanding a system's effective memory capacity.

Active Memory Expansion employs memory compression technology to transparently compress in-memory data, allowing more data to be placed into memory and thus expanding the memory capacity of POWER7 technology-based systems. Utilizing Active Memory Expansion can improve system utilization and increase a system's throughput.

There are multiple papers explaining in detail the functionality and performance of AME, for example:

- ▶ Active Memory Expansion: Overview and Usage Guide  
<ftp://ftp.software.ibm.com/common/ssi/sa/wh/n/pow03037usen/P0W03037USEN.PDF>
- ▶ Active Memory Expansion Performance  
<ftp://ftp.software.ibm.com/common/ssi/sa/wh/n/pow03038usen/P0W03038USEN.PDF>
- ▶ A proof of concept for SAP Retail and DB2, which also used AME  
<http://www.sdn.sap.com/irj/scn/go/portal/prtrroot/docs/library/uuid/b074b49a-a17f-2e10-8397-f2108ad4257f?QuickLink=index&overridelayout=true>

AME is supported for several SAP applications. Detailed information can be found in SAP Note 1464605 - *Active Memory Expansion (AME)*. This note also documents prerequisites and database support statements.

Various SAP applications have been successfully tested with AME and the test results were documented in the previously mentioned papers. It turned out that ABAP application servers can benefit quite well from AME and in some cases activating AME on the database will also show improvements in memory utilizations. Due to the access patterns, especially during garbage collection, SAP Application Server Java-based applications cannot take full advantage of AME and are not good candidates for AME.

**Tip:** If you want to get experience with AME but do not want to start right away, enable AME in the profile of the LPAR and set the expansion factor to 1.0. Then you can start testing AME at any time by changing the expansion factor with DLPAR operations without the need to reboot the LPAR.

For further details, see the white paper *Active Memory Expansion Performance* at:

<http://public.dhe.ibm.com/common/ssi/ecm/en/pow03038usen/P0W03038USEN.PDF>

## 8.6 Processor utilization metrics

Here we discuss technical details about measuring the utilization of processors in general, the effect of SMT on processor utilization metrics and the impact on SAP applications in particular. The reason is that new innovations in hardware are changing the way processor utilization and other metrics need to be interpreted.

### 8.6.1 Introduction

Processor time is a metric that quantifies the amount of processor resources a process used. Equivalently, processor utilization is the relation between the used and available resources. For example, the documentation of the `times()` system call in the UNIX standard at

<http://pubs.opengroup.org/onlinepubs/7990989775/xsh/times.html>

states that the field `tms_utime` returns the “processor time charged for the execution of user instructions of the calling process”. Intuitively one would assume this time to be equivalent to the time a process is active on the processor. Unfortunately, this is not exactly right and we describe in the following sections some background and effects on processor monitoring in general, and also for SAP specifically.

## 8.6.2 Test case

To illustrate some effects we used a small and simple test program written in C. Following is the source code if you would like to run some experiments on your own. We assume the source is named `test_cpu.c`:

```
#include <sys/times.h>
#include <unistd.h>
#include <stdio.h>

static clock_t start_time;
static clock_t end_time;
static struct tms start_cpu;
static struct tms end_cpu;
static unsigned long clock_ticks;
static double run_time = 0;
static double user_time = 0;
static double sys_time = 0;
static double cpu_time = 0;

int main( int argc, char ** argv )
{
    clock_ticks = sysconf(_SC_CLK_TCK );
    start_time = times( &start_cpu );

    while ( run_time < 10 )
    {
        end_time = times( &end_cpu );
        run_time = (double)(end_time - start_time) / clock_ticks;

        user_time = (double)( end_cpu.tms_utime - start_cpu.tms_utime) / clock_ticks;
        sys_time = (double)( end_cpu.tms_stime - start_cpu.tms_stime) / clock_ticks;
        cpu_time = user_time + sys_time;

        printf("Process           : %d\n", getpid());
        printf(" Ticks / sec       : %d\n", clock_ticks );
        printf(" Elapsed time      : %2.3f sec\n", run_time);
        printf(" Charged CPU time  : %2.3f sec\n", cpu_time );

        return 0;
    }
}
```

You can compile the program on AIX with

```
cc -o test_cpu test_cpu.c
```

or on Linux with

```
gcc -o test_cpu test_cpu.c
```



You can create a similar program on IBM i based on the API QUSRJOB1 with format JOB11000.

This program runs continuously in a loop and calls the times() API. This API returns an elapsed time since an arbitrary time in the past. By comparing the value taken at start-up of the program and continuously in the loop, the program can easily check how long it ran. After 10 seconds it stops. The times() API returns the charged processor time, too. At the end the program prints out the elapsed time and the charged processor time. As such, this program is an ideal test case to report the charged processor time for a given interval of time.

### 8.6.3 Processor time measurement and SMT

As described in the introduction, the UNIX standard does not define the exact semantics of processor time. Today different operating systems implement different semantics. In one case the processor time is reported as the time a process was active on the processor. In the other case it reports the charged processor time related to the used resources within a processor core.

#### Running the test program on a single core without SMT

In the simple case, that is, only one thread is running in a core (which is equivalent to SMT-1 on POWER), the situation is easy. The previously introduced test program shows the effect nicely.

If the program runs in an LPAR with one core and SMT-1, we are getting the following result (Example 1 in Figure 8-2 on page 88):

```
Process          : 360680
Ticks / sec      : 100
Elapsed time     : 10.000 sec
Charged CPU time : 9.970 sec
```

If we run two instances of the test program in parallel in the same LPAR, the output from the two processes looks like the following (Example 2 in Figure 8-2):

```

Process          : 327778
Ticks / sec     : 100
Elapsed time    : 10.010 sec
Charged CPU time : 4.980 sec

Process          : 360624
Ticks / sec     : 100
Elapsed time    : 10.000 sec
Charged CPU time : 4.990 sec
    
```

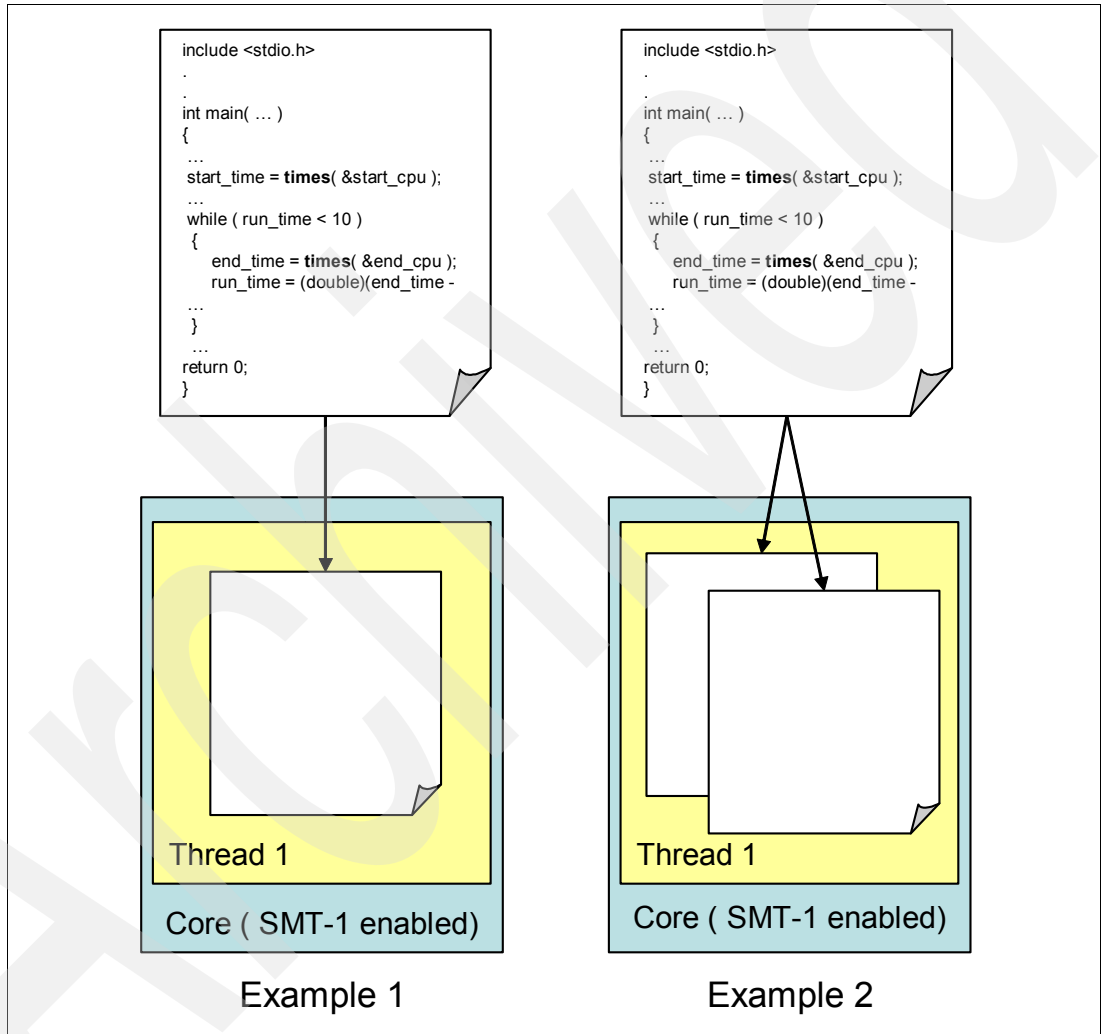


Figure 8-2 Running the test program on a single core without SMT

This result is easy to understand. There is only one core and one thread, so at any given time only one of the processes (here the process with process ID 327778 and 360624) can be active on the processor. Therefore, each process got charged for the time it was active on the processor. Independent of the operating system charging processor time based on resources (AIX) or on time active on the processor (Linux on x86), the results are the same.

For operating systems charging for used resources, the process had all resources available when it was on the processor, but it was only half of the time scheduled on the processor by

the operating system. Consequently, it is getting charged only for about 5 seconds of processor time.

In case of operating systems charging for time active on the processor, the operating system scheduled one process only half of the time and therefore also in this case 5 seconds of processor time are charged.

For the case with only one thread in a core, the reporting of processor time is simple, intuitive, and consistent even between operating systems using different accounting methods for processor time.

## **Fundamentals of SMT**

For several years now it is also possible to run multiple hardware threads on a single core. On POWER5 this was introduced as SMT and allowed two threads in one core. Starting with POWER7, up to four threads per core are possible. Other processors allow multiple threads too, for example the Hyper-Threading technology in several Intel processors. SMT is explained very well in multiple papers, for example:

[http://www.ibm.com/developerworks/aix/library/au-aix5\\_cpu/](http://www.ibm.com/developerworks/aix/library/au-aix5_cpu/)

Simply said, SMT provides mechanisms in a core to share the available resources between multiple threads, and thus achieve better utilization of existing resources within the core. Even though a single thread in a core looks to the operating system like a complete and independent processor, one has to remember that depending on the SMT mode multiple threads all share the same resources of the core and therefore may conflict with each other. As a result, SMT is increasing the throughput of a core but may decrease the response times of threads.

For operating systems such as AIX, charging the processor time based on used resources, SMT introduced some additional challenges. Since all threads use the same resources in the core, a mechanism had to be identified that would allow quantifying the usage of resources in a core on a thread basis. For this purpose POWER5 introduced the Processor Utilization Resource Register (PURR) in the processor. Each thread has its own register, which is incremented with a portion of the increment of the time base register, which is proportional to the amount of the used resources. The time base register is incremented monotonously with the time and can be used to represent the elapsed time. Therefore, the content of the PURR represents a portion of the elapsed time that is proportional to the used resources of a core.

More information about the PURR-based processor time accounting can be found at:

<http://www.ibm.com/developerworks/wikis/display/WikiPtype/Understanding+Processor+Utilization+on+POWER+Systems+-+AIX>

## **Running the test program on a single core with SMT**

Looking at the test program again we can directly see the effects of the different implementations for charging processor time.

The following examples run two instances of the program on a single core with two hardware threads enabled.

On AIX we see the following result (Example 3 in Figure 8-3):

```
Process          : 376978
Ticks / sec      : 100
Elapsed time     : 10.000 sec
Charged CPU time : 4.780 sec

Process          : 196642
Ticks / sec      : 100
Elapsed time     : 10.000 sec
Charged CPU time : 5.170 sec
```

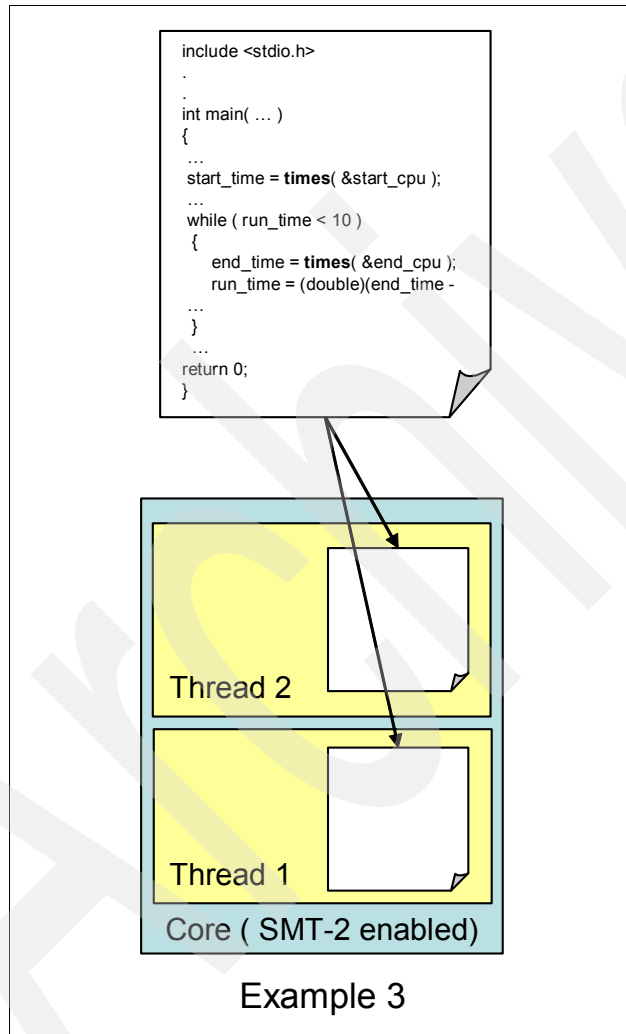


Figure 8-3 Running the test program on a single core with SMT

Both instances of the program are simultaneously active on the core, but each thread can use only parts of the resources during this time, since it has to share them with the other thread. As a result we see that each thread is only charged about 5 seconds processor time. In total roughly 10 seconds of processor time have been charged for 10 seconds elapsed time, which means 100% of time the core was used. In other words, the core was 100% utilized. We also see that the processor time of a thread does not relate in any form to the time the thread was active on the core.

Looking at the data produced by an operating system charging for time active on the processor we get very different results. The following output was produced on a Linux x86 system with one core and Hyper- Threading enabled:

```
Process          : 4673
Ticks / sec      : 100
Elapsed time     : 10.000 sec
Charged CPU time : 9.960 sec

Process          : 4672
Ticks / sec      : 100
Elapsed time     : 10.000 sec
Charged CPU time : 9.970 sec
```

Also in this case, both threads were active on the processor all the time. In contrast to AIX, each thread is charged for almost the whole elapsed time. So from the processor time it can be directly derived how long a thread was active on a core. On the other hand, no conclusion can be made on the amount of resources used. It cannot be determined that a single thread in the previous case had only half of the resources of the core available and therefore potentially had an impact on the work it was able to execute during this time.

**Summary:** The metric processor time is not exactly defined in the UNIX standard, and different operating systems provide different semantics. In one case it represents the active time on the processor whereas in the other case it represents an equivalent to the used resources. Both semantics provide very useful information. Processor time as active time on the processor can help in the analysis of application performance issues, for example to decide where a process spent its time. Processor time as a measurement for the used resources can help better to determine the overall utilization and can help to answer how much headroom in the sense of computing power is left on the system.

## 8.6.4 PURR-based metrics in POWER7

As described earlier, the increments in the time base register are proportionally distributed to the PURR registers in relation to the used resources of the threads. In POWER5 and POWER6 only threads doing active work got increments in their PURR registers. This lead to some artifacts in cases with only one active thread where the active thread was charged for all resources and a system seemed to be 100% utilized, even though the second thread would have been able to provide additional computing power.

With POWER7 this special case has been adjusted. If one or more threads are inactive, the active threads are not charged for all resources in the core. The inactive threads are also charged for those resources they potentially could use. The test program nicely illustrates the effects.

Result for one active instance of the program on one core with SMT-2 enabled (Example 4 in Figure 8-4 on page 92):

```
Process          : 286730
Ticks / sec      : 100
Elapsed time     : 10.000 sec
Charged CPU time : 8.190 sec
```

Result for one active instance of the program on one core with SMT-4 enabled (Example 5 in Figure 8-4):

```

Process           : 286734
Ticks / sec       : 100
Elapsed time      : 10.000 sec
Charged CPU time  : 6.250 sec
  
```

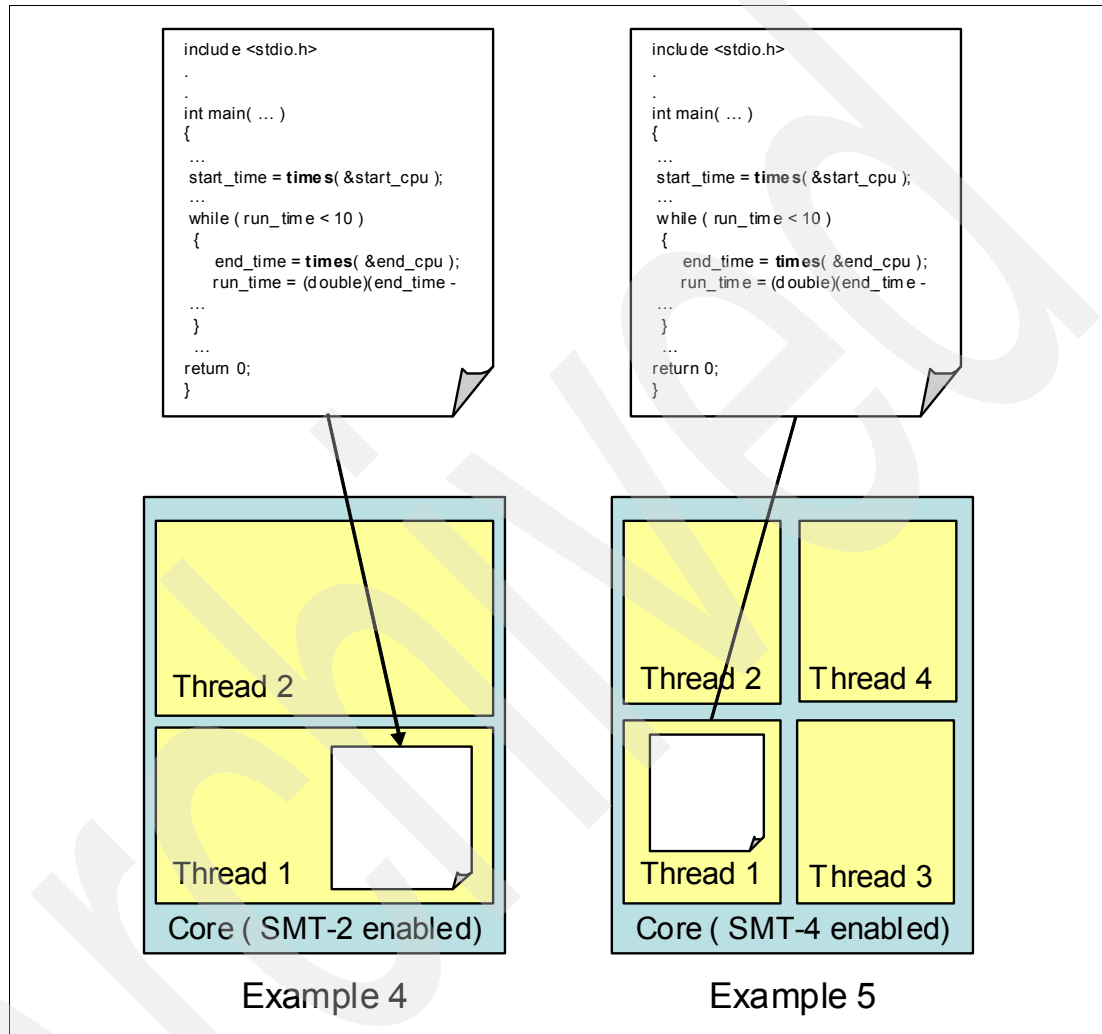


Figure 8-4 Running the test program on cores with SMT enabled

The charged processor time shows that there is still headroom available in the core to execute additional workload. The remaining headroom can only be used by additional processes scheduled on the free threads of the core, since the first process already uses one thread of the core all the time.

### 8.6.5 Processor time in SAP applications

In SAP applications the processor time is also used in several places, most importantly in the single statistic records, which are written for every dialog step and report a number of metrics gathered during the dialog step. For example, the records contain the time spent in the database, in the GUI and in other areas. The processor time is also returned as one of the metrics. The single statistic records can be shown with the transaction STAD. In the SAP

system this very granular information is also aggregated and used for reporting in transaction ST03.

### Using single statistic records for application performance analysis

The single statistic records or the derived results in ST03 are used for multiple purposes. One of them is the analysis of application performance issues. There is a rule of thumb that was established several years ago and not changed since then. This rule says: "If the Processing Time is two times higher (or more) than the processor time, then we likely have a processor bottleneck". While this rule was valid in the early days, where systems did not have the capability of hardware threads (SMT), it is not applicable any more on systems reporting the processor time as resource usage. Figure 8-5 shows a single statistic record for an ABAP program which just ran on the processor and had no activities on the database, and so on. The LPAR was running in SMT-4 mode.

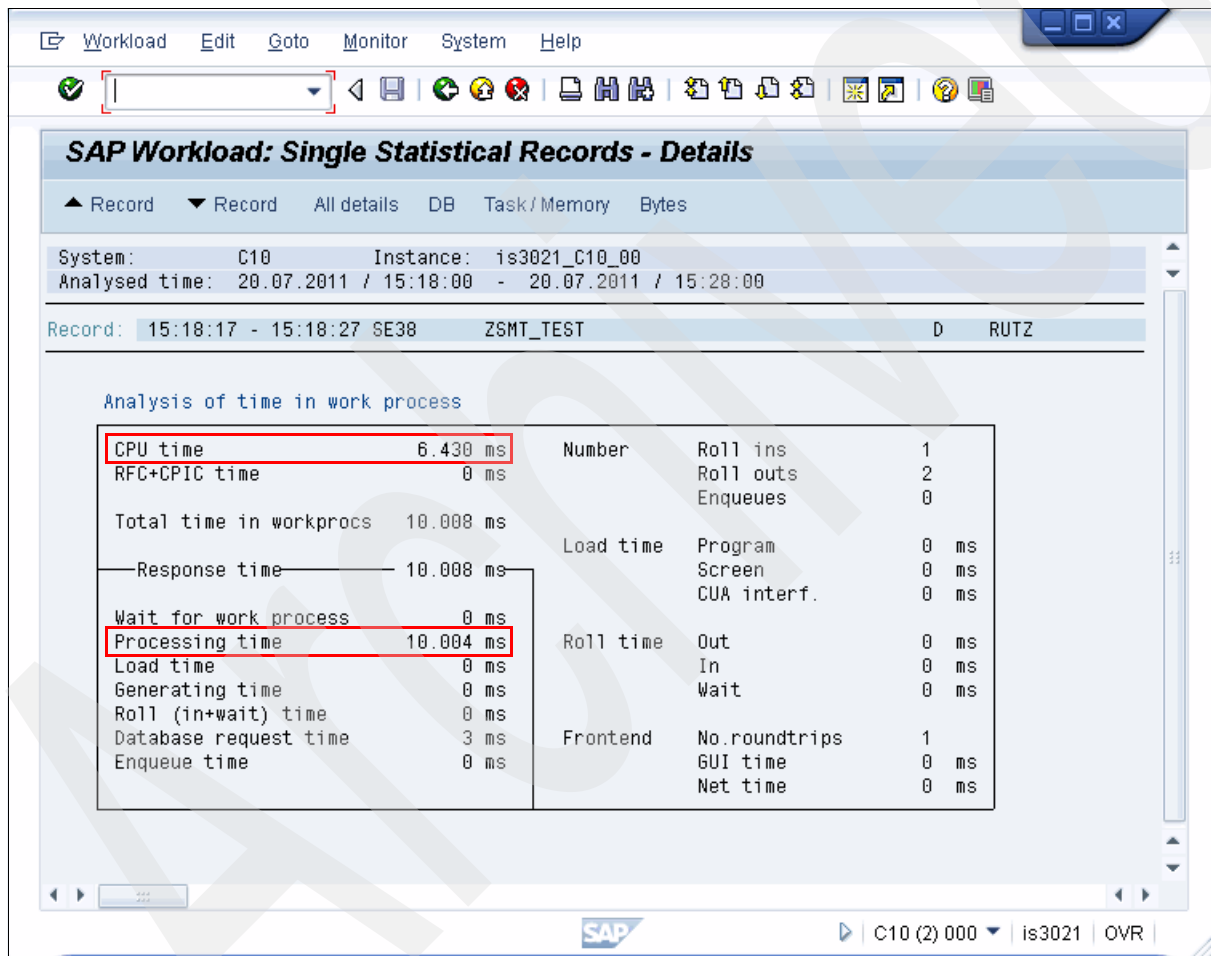


Figure 8-5 Detailed view of a single statistics record

The processor time reported in this screen shot is the processor time reported from the operating system.

We see the same behavior for this APAP program as we have shown earlier with our test program. Even though the program was active for the whole 10 seconds, it got charged only for 6.43 seconds, which indicates that there is free headroom in the core for additional work. The previously mentioned rule of thumb would now indicate that we are moving towards a processor bottleneck, which is definitively not the case.

**Important:** Some rules of thumb for the analysis of application performance issues in SAP systems based on single statistic records do not apply on POWER systems when SMT is enabled.

### Using single statistic records for workload sizing

The other use of single statistic records is to determine the amount of computing power required for certain applications in order to size systems. The processor times for the transactions related to the application would be added up and set into relation of the measurement time interval. In this case the single statistic records on Power Systems provide with the reported processor time a good indicator for proper sizing of workload. On systems reporting processor time as time active on the processor this method returns inaccurate information for the required resources to run the measured workload.

**Important:** The single statistic records in SAP CCMS and their derived metrics in ST03 are providing a good set of data for application sizing purposes on POWER systems.

## 8.6.6 Side effects of PURR based metrics in Shared Pool environments

AIX partitions in a shared processor pool environment provide two metrics describing the usage of processors, physb and physc. The metric physc refers to the amount of computing power in units of processors provided by the hypervisor to a shared processor LPAR. This metric does not take into account how many of the threads were actually used by the LPAR, since the hypervisor can assign only full cores to an LPAR. In contrast to this the metric physb describes the amount of computing power used by the LPAR.

With the changes in the PURR handling in POWER7 technology-based systems, in some cases more idle time is reported to correctly represent the available free headroom in the system. As a result the difference between physb and physc may get larger, and also some existing recommendations have to be adjusted.

Since POWER7 technology-based systems provide four threads per core, it is not a rare case that physb is showing much lower values than physc and as such indicating free headroom within the resources provided by the hypervisor. This situation is also promoted by the fact that the operating system primarily schedules workload to the primary threads in the available cores, before the secondary and tertiary threads are used. Monitoring only physc to determine the required computing power for an LPAR may lead to wrong conclusions, since the potential free resources within an LPAR are not taken into account.

In the SAP transaction ST06 the metric physc is shown as Capacity Consumed. The value for physb is shown as percentage of the Available Capacity. It is named Available Capacity Busy. The original value for physb can be calculated by multiplying Available Capacity with Available Capacity Busy.

The described changes result in slightly changed recommendations. If the value of physc is nearing the number of virtual processors in the system, the number of virtual processors should not immediately be increased, unless the value for physb also indicates that all resources are used. If physb is significantly lower than physc, the number of virtual processors could stay unchanged or may even be reduced. In all cases the specific requirements of the workload have to be taken into account.





## Monitoring

In this chapter, we mention some new and reintroduce some commonly used command line operating system (OS) tools that can monitor various virtualization features or aspects. The first part covers performance monitoring on AIX, the second covers IBM i. Linux running on Power Systems use SAP-based monitoring in CCMS with ST06 or open source tools such as Ganglia.

## 9.1 Performance Monitoring in AIX

We now describe how to integrate the virtual landscape into the traditional OS monitoring tools. We also introduce new tools that you can use to monitor different aspects of the virtual landscape. In order to improve readability, some columns of the output of commands may have been removed.

### 9.1.1 The vmstat monitoring tool

vmstat is a familiar monitoring tool that is found on most UNIX variants. It reports statistics about kernel threads, virtual memory, disks, traps, and processor activity. For system processor and memory utilization investigations, vmstat is a good starting point. With the introduction of shared processor partitions and micropartitions, vmstat was enhanced on AIX to include some new virtualization-related statistics.

For micropartitions, vmstat now reports the physical processor consumption by the LPAR and the percentage of entitlement that is consumed, which is identified by the pc and ec columns, respectively. The presence of the ec column also indirectly identifies the LPAR as a Shared Processor LPAR, as shown in Example 9-1.

*Example 9-1 vmstat report*

```
$ vmstat -w 1 5
```

```
System configuration: lprocessor=2 mem=1024MB ent=0.20
```

kthr		memory	page	faults	processor								
r	b	avm	fre	...	in	sy	cs	us	sy	id	wa	pc	ec
2	0	112808	126518	...	20	60	143	99	0	0	0	0.96	480.6
2	0	112808	126518	...	22	28	141	99	1	1	0	0.47	236.6
1	0	112808	126518	...	21	20	147	0	2	98	0	0.01	3.7
1	0	112808	126518	...	16	41	144	91	2	7	0	0.19	94.1
1	0	112808	126518	...	14	25	152	99	1	1	0	0.43	216.8

**Tip:** The `-w` flag of vmstat provides a formatted wide view, which makes reading vmstat output easier, given the additional columns.

From the WPAR global environment, `vmstat` can also report information about the WPAR level with the `-@` option, `-@ ALL` for example. Example 9-2 shows a summary line for the overall system and individual statistics for the global environment and each active WPAR.

*Example 9-2 vmstat report*

```
#vmstat -w -@ ALL 5 1
```

System configuration: lprocessor=2 mem=2048MB drives=0 ent=1.00 wpar=2

wpar	kthr		memory		page		faults				processor			
	r	b	avm	fre	...	in	sy	cs	us	sy	id	wa	pc	rc
System	1	0	223635	250681	...	16	31	187	77	0	23	0	0.87	87.1
Global	1	0	-	-	...	-	-	172	99	1	-	-	0.54	53.6
is3042w	1	0	-	-	...	-	-	15	99	0	-	-	0.24	23.7
is3043w	0	0	-	-	...	-	-	0	38	62	-	-	0.00	0.0

With the introduction of Active Memory Expansion the `vmstat` command has been enhanced with additional metrics that can be monitored. For more details about how AME works, refer to 8.5, “Active Memory Expansion for SAP systems” on page 84. The AME-related metrics can be monitored by adding the option `-c`, as shown in Example 9-3.

*Example 9-3 vmstat report including AME information*

```
#vmstat -w -c 1
```

System Configuration: lprocessor=8 mem=6144MB tmem=4096MB ent=0.20 mmode=dedicated-E

kthr		memory				page						...
r	b	avm	fre	csz	cfr	dxm	ci	co	pi	po	...	
1	0	422566	598766	4096	4096	0	0	0	0	0	...	
1	0	422566	598766	4096	4096	0	0	0	0	0	...	
1	0	422566	598766	4096	4096	0	0	0	0	0	...	
0	0	422621	598711	4096	4096	0	0	0	0	0	...	
1	0	422624	598708	4096	4096	0	0	0	0	0	...	
1	0	422566	598766	4096	4096	0	0	0	0	0	...	
0	0	422588	598744	4096	4096	0	0	0	0	0	...	

The size of the compressed pool is shown in column `csz` and the amount of free memory in the compressed pool as `cfr`. Similarly to the paging rates (`pi` and `po`), the rates for compressing and decompressing are shown as `ci` and `co`. Those values can be higher than the values acceptable for the paging rates, since compression and decompression are much faster than paging. The column `dxm` describes the memory deficit. If this value does not equal zero, the operating system cannot achieve the requested expansion rate.

### 9.1.2 The sar monitoring tool

Like `vmstat`, `sar` is a generic UNIX tool for collecting statistics about processor, I/O, and other system activities. You can use the `sar` tool to show both history and real-time data.

Beginning with AIX 5.3, the `sar` command reports virtualization-related processor utilization statistics. For micropartitions, the physical processors consumed and the entitlement consumed are now shown, as in Example 9-4.

*Example 9-4 sar command report*

---

```
#sar 1 5

AIX is3013 3 5 00C46F8D4C00    09/09/08

System configuration: lprocessor=2 ent=0.20 mode=Uncapped

15:17:04   %usr   %sys   %wio   %idle   physc   %entc
15:17:05     100     0     0     0     1.00   499.7
15:17:06     100     0     0     0     1.00   499.7
15:17:07     99      1     0     0     0.57   284.9
15:17:08     99      1     0     0     0.61   306.2
15:17:09      4     54     0     42     0.01    2.8

Average     99      0     0     0     0.64   318.0
```

---

You can also monitor the metrics per processor, as shown in Example 9-5.

*Example 9-5 Metrics per-processor report*

---

```
#sar -P ALL 1 5

AIX is3013 3 5 00C46F8D4C00    09/09/08

System configuration: lprocessor=2 ent=0.20 mode=Uncapped

15:36:26 processor   %usr   %sys   %wio   %idle   physc   %entc
15:36:27  0      100     0     0     0     0.86   430.0
           1       1     8     0     92     0.00    0.5
           -      99     0     0     0     0.86   430.4
15:36:28  0       5    61     0     35     0.01    2.7
           1       2    21     0     77     0.00    0.4
           -       4    56     0     40     0.01    3.1
15:36:29  0      98     1     0     0     0.29   145.1
           1       0    16     0    84     0.00    0.4
           -      98     1     0     1     0.29   145.5
15:36:30  0     100     0     0     0     1.00   499.2
           1       1     7     0    92     0.00    0.5
           -     100     0     0     0     1.00   499.7
15:36:31  0     100     0     0     0     1.00   499.1
           1       1     7     0    92     0.00    0.5
           -     100     0     0     0     1.00   499.6
Average   0      99     0     0     0     0.63   315.2
           1       1    11     0    88     0.00    0.4
           -      99     0     0     0     0.63   315.7
```

---

### 9.1.3 The mpstat tool

The **mpstat** command collects and displays detailed output about performance statistics for all logical processors in the system. It presents the same metrics that **sar** does, but it also provides information about the run queue, page faults, interrupts, and context switches.

The default output from the **mpstat** command displays two sections of statistics, as shown in Example 9-6. The first section, which displays the system configuration, is shown when the command starts and whenever the system configuration is changed:

- ▶ **lprocessor**: The number of logical processors.
- ▶ **ent**: Entitled processing capacity in processor units. This information is displayed only if the partition type is shared.

The second section displays the utilization statistics for all logical processors. The **mpstat** command also displays a special processor row with the processor ID **ALL**, which shows the partition-wide utilization. The **mpstat** command gives the various statistics. It depends on the flag. The utilization statistics are:

- ▶ **processor**: Logical processor ID
- ▶ **min, maj**: Minor and major page faults
- ▶ **mpc**: Total number of inter-processor calls
- ▶ **int**: Total number of interrupts
- ▶ **cs, ics**: Total number of voluntary and involuntary context
- ▶ **rq**: Run queue size
- ▶ **mig**: Total number of thread migrations to another logical processor
- ▶ **lpa**: Number of re-dispatches within affinity domain 3
- ▶ **sysc**: Total number of system calls
- ▶ **us, sy, wa, id**: Processor usage statistics
- ▶ **pc**: The percentage of entitlement consumed
- ▶ **%ec**: Fraction of processor consumed
- ▶ **lcs**: Total number of logical context switches

*Example 9-6 mpstat command report*

```
#mpstat 1 2

System configuration: lprocessor=8 ent=1.0 mode=Uncapped

processor min maj mpc int cs ics rq mig lpa sysc us sy wa id pc %ec lcs
0 19 0 0 117 46 20 0 2 100 128 98 2 0 0 0.18 8.8 120
1 0 0 0 101 27 13 0 2 100 4 100 0 0 0 0.55 27.3 103
2 0 0 0 265 124 66 0 0 100 1 100 0 0 0 0.80 39.4 114
3 0 0 0 102 0 0 0 0 - 0 100 0 0 0 0.20 10.0 113
4 0 0 0 93 1 1 0 1 100 0 100 0 0 0 0.29 14.4 98
5 0 0 0 88 0 0 0 0 - 0 0 36 0 64 0.00 0.0 90
6 0 0 0 88 0 0 0 0 - 0 0 32 0 68 0.00 0.1 88
7 0 0 0 88 0 0 0 0 - 0 0 36 0 64 0.00 0.1 88
ALL 19 0 0 942 198 100 0 5 100 133 99 0 0 0 2.02 202.3 814

-----
0 3 0 0 132 41 18 1 0 100 59 12 52 0 36 0.00 0.1 120
1 0 0 0 107 27 13 0 0 100 5 100 0 0 0 0.91 31.2 107
2 0 0 0 268 125 66 1 0 100 0 100 0 0 0 1.00 34.3 110
3 0 0 0 104 0 0 0 0 - 0 0 34 0 66 0.00 0.0 105
4 0 0 0 112 3 3 1 0 100 0 100 0 0 0 1.00 34.3 107
5 0 0 0 91 0 0 0 0 - 0 0 22 0 78 0.00 0.0 91
6 0 0 0 92 0 0 0 0 - 0 0 36 0 64 0.00 0.0 90
7 0 0 0 90 0 0 0 0 - 0 0 38 0 62 0.00 0.0 90
```

```
ALL 3 0 0 996 196 100 3 0 100 64 100 0 0 0 2.91 291.0 820
```

---

If the partition type is shared, a special processor row with the processor ID U is displayed when the entitled processing capacity is not entirely consumed.

With the option **-s** the **mpstat** command provides information about the utilization of each logical processor as well as the utilization of a complete core, as you see in Example 9-7. It can also be determined which logical processor belongs to a certain core.

*Example 9-7 mpstat command information*

---

```
#mpstat -w -s 1 2
```

System configuration: lprocessor=8 ent=0.2 mode=Uncapped

Proc2		Proc0					
52.83%		28.13%					
processor0	processor1	processor4	processor5	processor2	processor3	processor6	processor7
33.57%	6.02%	6.79%	6.45%	19.62%	3.50%	2.51%	2.51%
-----							
Proc2		Proc0					
20.18%		0.28%					
processor0	processor1	processor4	processor5	processor2	processor3	processor6	processor7
13.92%	2.54%	1.68%	2.03%	0.10%	0.06%	0.07%	0.06%

---

In Example 9-7 the logical processors 0, 1, 4 and 5 (which map to the threads in one core) belong to processor 2 which is equivalent to a core. This view of **mpstat** helps to understand how workload is dispatched to the threads and cores and how processor folding may influence the system at a given time.

The **mpstat** command can also provide a lot more information when you use additional parameter flags.

### 9.1.4 The lparstat tool

The **lparstat** command, introduced in AIX 5.3, is a useful tool for displaying the LPAR configuration. You can get information about the LPAR configuration with the **-i** option, as shown in Example 9-8.

*Example 9-8 lparstat command information*

---

```
#lparstat -i
Node Name           : is3046
Partition Name      : is3046
Partition Number    : 46
Type                : Shared-SMT-4
Mode                : Uncapped
Entitled Capacity   : 0.20
Partition Group-ID  : 32814
Shared Pool ID      : 0
Online Virtual processors : 2
Maximum Virtual processors : 8
Minimum Virtual processors : 1
Online Memory       : 4096 MB
Maximum Memory      : 16384 MB
```

```

Minimum Memory                : 512 MB
Variable Capacity Weight      : 128
Minimum Capacity              : 0.10
Maximum Capacity              : 0.80
Capacity Increment            : 0.01
Maximum Physical processors in system : 32
Active Physical processors in system : 32
Active processors in Pool     : 26
Shared Physical processors in system : 26
Maximum Capacity of Pool     : 2600
Entitled Capacity of Pool    : 1820
Unallocated Capacity         : 0.00
Physical processor Percentage : 10.00%
Unallocated Weight           : 0
Memory Mode                   : Dedicated-Expanded
Total I/O Memory Entitlement  : -
Variable Memory Capacity Weight : -
Memory Pool ID                : -
Physical Memory in the Pool   : -
Hypervisor Page Size         : -
Unallocated Variable Memory Capacity Weight: -
Unallocated I/O Memory entitlement : -
Memory Group ID of LPAR      : -
Desired Virtual processors    : 2
Desired Memory                : 4096 MB
Desired Variable Capacity Weight : 128
Desired Capacity              : 0.20
Target Memory Expansion Factor : 1.50
Target Memory Expansion Size  : 6144 MB
Power Saving Mode             : Disabled

```

---

Additionally, detailed utilization metrics can also be displayed, as shown in Example 9-9. The default output contains two sections: a configuration section and a utilization section.

*Example 9-9 mpstat command output*

---

```
#lparstat 1 10
```

```
System configuration: type=Shared mode=Uncapped smt=4 lprocessor=8 mem=6144MB
psize=26 ent=0.20
```

%user	%sys	%wait	%idle	phyc	%entc	lbusy	app	vcs	phint
17.9	45.7	0.0	36.4	0.48	237.6	6.0	25.01	293	0
18.7	47.9	0.0	33.4	0.55	275.1	7.8	24.94	426	3
18.9	47.1	0.0	34.1	1.00	498.4	13.4	23.82	360	0
1.9	10.5	0.0	87.6	0.07	35.6	0.3	24.54	777	0
18.8	47.5	0.0	33.7	0.98	491.2	16.7	23.78	392	24
17.8	45.4	0.0	36.7	0.61	303.3	7.1	23.90	334	13
18.5	48.5	0.0	33.0	0.44	219.2	6.2	23.72	655	12

---

Here are definitions of the metrics:

- %user, %sys: The percentage of the physical processor used while executing at the user and system levels

- ▶ %wait, %idle: I/O wait and idle levels
- ▶ physc: The number of physical processors that are consumed
- ▶ %entc: The percentage of the entitled capacity consumed
- ▶ lbusy: The percentage of logical processor(s) utilization that occurred while executing at the user and system level
- ▶ App: The available physical processors in the shared pool
- ▶ Vcsw: The number of virtual context switches that are virtual processor hardware preemptions
- ▶ Phint: The number of phantom (targeted to another shared partition in this pool) interruptions that are received

Equivalent to the **vmstat** command, **lparstat** can provide AME metrics with the option **-c**, as in Example 9-10.

*Example 9-10 lparstat command output*

```
#lparstat -c 1 10
```

```
System configuration: type=Shared mode=Uncapped mmode=Ded-E smt=4 lprocessor=8
mem=6144MB tmem=4096MB psize=26 ent=0.20
```

%user	%sys	%wait	%idle	physc	%entc	lbusy	app	vcsw	phint	%xprocessor	xphysc	dxm
18.2	45.7	0.0	36.1	0.84	417.8	10.9	24.15	380	3	0.0	0.0000	0
18.8	50.8	0.0	30.4	0.21	103.3	3.4	25.33	512	0	0.0	0.0000	0
18.7	46.7	0.0	34.6	1.00	500.5	16.7	24.55	386	6	0.0	0.0000	0
17.1	44.6	0.0	38.3	0.41	202.8	5.2	25.16	302	0	0.0	0.0000	0
18.7	48.2	0.0	33.1	0.62	312.3	9.1	24.06	421	2	0.0	0.0000	0
18.3	45.9	0.0	35.8	0.96	479.8	12.5	24.54	388	2	0.0	0.0000	0
7.1	22.3	0.0	70.6	0.08	41.1	1.1	25.43	544	0	0.0	0.0000	0
18.5	46.1	0.0	35.3	1.01	506.5	13.0	24.54	258	1	0.0	0.0000	0
18.1	46.1	0.0	35.9	0.51	255.9	8.7	25.05	429	0	0.0	0.0000	0
18.9	48.5	0.0	32.7	0.50	252.0	7.4	24.07	436	1	0.0	0.0000	0

As in **vmstat**, the deficit memory is shown in column **dxm**. Additionally, the amount of processor capacity, which is required for compression and decompression, is documented in columns **%xprocessor** and **xphysc**.



As described in 8.6, "Processor utilization metrics" on page 85, the reporting of processor utilization changed with POWER7 technology-based systems in some cases. There it was also outlined, that in shared pool environments the metrics physc and physb can show a larger difference than in the past. If you are interested in the computing power an LPAR used (physb), the **lparstat** command can provide this information as well as part of the metrics reported with the option **-m**, as in Example 9-11.

*Example 9-11 lparstat -m command output*

```
#lparstat -m 1 10
```

---

System configuration: lprocessor=8 mem=6144MB mpsz=0.00GB iome=4096.00MB iomp=10 ent=0.20

physb	hpi	hpit	pmem	iomin	iomu	iomf	iohwm	iomaf	%entc	vcsw
30.42	0	0	4.00	48.2	12.1	-	26.7	0	240.3	1013
84.17	0	0	4.00	48.2	12.1	-	26.7	0	620.2	877
44.64	0	0	4.00	48.2	12.1	-	26.7	0	349.7	607
22.82	0	0	4.00	48.2	12.1	-	26.7	0	174.7	709
65.71	0	0	4.00	48.2	12.1	-	26.7	0	499.6	464
16.51	0	0	4.00	48.2	12.1	-	26.7	0	135.4	658
50.84	0	0	4.00	48.2	12.1	-	26.7	0	386.4	499
52.74	0	0	4.00	48.2	12.1	-	26.7	0	407.7	274
14.27	0	0	4.00	48.2	12.1	-	26.7	0	109.0	514
65.62	0	0	4.00	48.2	12.1	-	26.7	0	494.4	375

---

## 9.1.5 The topas tool

Many AIX system administrators and analysts use the **topas** command to view general statistics about the local system activity. The data is displayed in a simple and convenient character-based format using the curses library.

The **topas** tool was expanded to include LPAR information. Start **topas** with the **-L** option or press **L** while **topas** is running and panel output, similar to Example 9-12, is displayed.

*Example 9-12 topas -L*

---

```
Interval: 2 Logical Partition: is3051 Mon Oct 6 14:36:32 2008
Psize: 15 Shared SMT ON Online Memory: 2048.0
Ent: 0.40 Mode: UnCapped Online Logical processors: 8
Partition processor Utilization Online Virtual
processors: 4
```

%usr	%sys	%wait	%idle	physc	%entc	%lbusy	app	vcsw	phint	%hypv	hcalls
100	0	0	0	2.6	658.46	34.06	10.70	681	15	26.2	509

---

Lprocessor	minpf	majpf	intr	csw	icsw	runq	lpa	scalls	usr	sys	_wt	idl	pc	lcsw
processor0	0	0	283	134	65	0	100	36	99	0	0	0	0.63	146
processor1	0	0	103	2	2	0	100	0	100	0	0	0	0.18	102
processor2	0	0	107	4	4	0	100	3	100	0	0	0	0.38	101
processor3	0	0	101	0	0	0	100	0	91	3	0	6	0.01	101
processor4	0	0	111	9	9	0	100	1	100	0	0	0	1.00	105
processor5	0	0	10	0	0	0	0	0	0	8	0	92	0.00	10
processor6	0	0	106	5	4	0	100	5	100	0	0	0	0.44	102
processor7	0	0	11	0	0	0	0	0	0	15	0	85	0.00	11

---

The upper section of Example 9-12 on page 103 shows a subset of the `lparstat` command statistics, while the lower section displays a subset of `mpstat` data.

The tool also provides a useful cross partition view that displays metrics from all AIX (5.3 TL3 or later) and VIOS (1.3 or later) partitions on the same host server, as shown in Example 9-13. On the command line you can get directly to this panel with the option `-C`.

Example 9-13 Cross partition view, `topas -C`

```

Topas CEC Monitor                Interval: 10                Mon Oct  6 14:54:01 2008
Partitions Memory (GB)          Processors
Shr: 10  Mon:89.0 InUse:68.1  Shr:2.6  PSz: -  Don: 0.0 Shr_PhysB 0.15
Ded:  3  Avl:  -                Ded:  5  APP: -  Stl: 0.0 Ded_PhysB 0.98

Host      OS  M Mem InU Lp  Us Sy Wa Id  PhysB  Vcsw Ent  %EntC PhI
-----shared-----
is3013    A53 U 1.0 0.9 4   0 22 0 77  0.05 169 0.20 23.3  2
is3053    A61 U 8.0 5.0 4   1  4 0 94  0.02 481 0.20  7.9  2
is17d4    A53 U 8.0 3.9 4   4  2 0 93  0.02 263 0.20  7.9  0
is32d1    A53 U 8.0 8.0 4   2  3 0 94  0.01 253 0.20  6.9  2
is308v2   A53 U 1.0 1.0 4   0  2 0 97  0.01 408 0.40  3.0  0
is20d1    A53 U 8.0 8.0 4   1  2 0 95  0.01 289 0.20  5.8  0
is33d1    A53 U 8.0 6.0 4   0  2 0 96  0.01 205 0.20  4.5  0
is308v1   A53 U 1.0 1.0 4   0  1 0 98  0.01 225 0.40  2.0  1
is25d1    A53 U 12 12 4   0  1 0 97  0.01 192 0.20  3.7  0
is3036    A53 U 8.0 5.2 8   0  0 0 99  0.01 201 0.40  1.5  0

Host      OS  M Mem InU Lp  Us Sy Wa Id  PhysB  Vcsw %istl %bstl %bdon %idon
-----dedicated-----
is3024    A61 D 8.0 4.7 4  49 0 0 49  0.98 476  0.00 0.00 0.00 0.07
is3025    A61 S 10 9.0 2   0  0 0 99  0.00 344  0.00 0.00  -  -
is3060    A53 S 8.0 3.6 4   0  0 0 99  0.00 303  0.00 0.00  -  -

```

This CEC view has also been enhanced with metrics for the throughput of the Virtual I/O Servers, as in Example 9-14. When the CEC view is on the window, just press `v` and you are getting to the VIO Server/Client section.

Example 9-14 `topas -L`

```

Topas VIO Server/Client Throughput Details host:is3025  Interval: 10
Vio Servers: 2          Vio clients: 26                Mon Jul 18 13:32:49 2011
-----
Server    KBPS    TPS    KB-R    ART    MRT    KB-W    AWT    MWT    AQW    AQD
is314v1   0.0     0.0    0.0     0.0    0.0    0.0     0.0    0.0    0.0    0.0
is314v2   0.0     0.0    0.0     0.0    0.0    0.0     0.0    0.0    0.0    0.0
-----
Client    KBPS    TPS    KB-R    ART    MRT    KB-W    AWT    MWT    AQW    AQD
is20d5    0.0     0.0    0.0     0.0    105.7K 0.0     0.0    47.6K 0.0    0.0
is27d2    0.0     0.0    0.0     0.0    104.0   0.0     0.0    91.0   0.0    0.0
is20d4    0.0     0.0    0.0     0.0    567.0   0.0     0.0    104.6K 0.0    0.0
is19d3    88.0    9.0     0.0     0.0    140.0   9.0     0.0    17.4K 0.0    0.0
is3025    0.0     0.0    0.0     0.0    3.3K    0.0     0.0    7.8K 0.0    0.0
is28d1    0.0     0.0    0.0     0.0    757.0   0.0     0.0    53.7K 0.0    0.0
is34d1    0.0     0.0    0.0     0.0    34.9K   0.0     0.0    5.2K 0.0    0.0
is30d1    0.0     0.0    0.0     0.0    2.0K    0.0     0.0    27.2K 0.0    0.0
is24d2    0.0     0.0    0.0     0.0    10.8K   0.0     0.0    5.3K 0.0    0.0
is31d1    0.0     0.0    0.0     0.0    701.0   0.0     0.0    1.8K 0.0    0.0

```

is29d2	48.0	3.0	0.0	0.0	342.0	3.0	0.0	31.3K	0.0	0.0
is3106	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
is3065	0.0	0.0	0.0	0.0	12.8K	0.0	0.0	655.0	0.0	0.0
is3093	0.0	0.0	0.0	0.0	1.1K	0.0	0.0	238.0	0.0	0.0
is3068	0.0	0.0	0.0	0.0	47.0	0.0	0.0	1.3K	0.0	0.0
is3092	8.0	2.0	0.0	0.0	1.0K	2.0	0.0	18.4K	0.0	0.0
is3063	0.0	0.0	0.0	0.0	37.8K	0.0	0.0	596.0	0.0	0.0
is3096	0.0	0.0	0.0	0.0	46.8K	0.0	0.0	5.6K	0.0	0.0
is3011	152.0	37.0	0.0	0.0	39.5K	37.0	0.0	3.4K	0.0	0.0
is3098	1.9Ko	113.3	0.5	0.0	11.7K	112.9	0.0	4.3K	0.0	0.0
is3100	0.0	0.0	0.0	0.0	1.4K	0.0	0.0	1.2K	0.0	0.0
is3102	0.0	0.0	0.0	0.0	11.6K	0.0	0.0	274.0	0.0	0.0
is3109	0.0	0.0	0.0	0.0	240.0	0.0	0.0	463.0	0.0	0.0
is3112	1.5Ko	49.0	0.0	0.0	15.9K	49.0	0.0	351.0	0.0	0.0
is3062	0.0	0.0	0.0	0.0	12.9K	0.0	0.0	1.6K	0.0	0.0

With larger systems and growing memory the memory topology is getting more important. Topas is providing the additional option **-M** to show the actual topology and the memory usage, as shown in Example 9-15.

*Example 9-15 topas -M output*

```

TopasMonitor for host: is3046 Interval: 2 Fri Jul 15 08:28:17 2011
=====
REF1 SRAD TOTALMEM INUSE FREE FILECACHE HOMETHRDS processorS
-----
0 0 3948.2 3674.4 273.9 2116.9 403 0-7
=====
processor SRAD TOTALDISP LOCALDISP% NEARDISP% FARDISP%
-----
0 0 394 100.0 0.0 0.0
1 0 77 100.0 0.0 0.0
2 0 7 100.0 0.0 0.0
3 0 0 0.0 0.0 0.0
4 0 0 0.0 0.0 0.0
7 0 0 0.0 0.0 0.0
5 0 0 0.0 0.0 0.0
6 0 0 0.0 0.0 0.0

```

Also AME metrics can now be seen at the main panel of **topas** in the section AME, as in Example 9-16.

*Example 9-16 topas output*

```

Topas Monitor for host: is3046          EVENTS/QUEUES  FILE/TTY
Fri Jul 15 08:29:29 2011 Interval: 2   Cswitch  1677  Readch  3738.9K
                                           Syscall  771.5K Writech  395.9K
Kernel  49.4 |#####| Reads    2427 Rawin    0
User    25.1 |#####| Writes   1125 Ttyout   349
Wait    0.0 |#| Forks    206  Igets    0
Idle    25.6 |#####| Execs    154  Namei    4380
Physc = 0.88                               %Entc= 438.1 Runqueue  1.5  Dirblk   0
                                           Waitqueue 0.0
Network KBPS  I-Pack  O-Pack  KB-In  KB-Out
Total   57.9  185.3  98.7   36.2   21.7
                                           PAGING
                                           Faults  26288 % Comp   27
                                           MEMORY
                                           Real,MB 6144

```

Disk	Busy%	KBPS	TPS	KB-Read	KB-Writ	Steals	0	% Noncomp	34
Total	0.0	82.2	20.0	0.0	82.2	PgspIn	0	% Client	34
						PgspOut	0		
FileSystem		KBPS	TPS	KB-Read	KB-Writ	PageIn	0	PAGING SPACE	
Total		3.3K	1.3K	3.3K	0.1	PageOut	0	Size,MB	5184
						Sios	0	% Used	0
Name	PID	processor%	PgSp	Owner				% Free	
100									
burn_hal	54264070	41.6	0.4	root		AME			
vmmd	458766	0.9	1.2	root		TMEM,MB	4096	WPAR Activ	0
Xvnc	9306302	0.4	3.2	root		CMEM,MB	16	WPAR Total	0
kcawd	6619320	0.2	12.0	root		EF[T/A]	1.5/1.5	Press: "h"-help	
xmtopasa	8126488	0.1	2.2	root		CI:0.0	CO:0.0	"q"-quit	
xterm	55640472	0.1	8.6	root					
topas	14745716	0.1	1.8	root					
random	5046450	0.1	0.4	root					
xmgc	917532	0.0	0.4	root					
xmtopas	5439546	0.0	4.7	root					
gil	1245222	0.0	0.9	root					
mysqld	6357154	0.0	42.4	mysql					
icewm-de	8650828	0.0	1.1	root					
nfsd	3014782	0.0	1.8	root					
init	1	0.0	0.7	root					
java	8388836	0.0	42.4	tomcat					
kuxagent	6881304	0.0	37.9	root					
rmcd	9502788	0.0	6.5	root					
kulagent	4325578	0.0	7.7	root					
sendmail	4063258	0.0	1.1	root					

Many more examples of screenshots and explanations can be found on:

<https://www.ibm.com/developerworks/wikis/display/WikiPtype/topas>

Similarly, as with **nmon**, the **topasrec** command provides the capability to record certain metrics over time for an LPAR or even a whole system with the option **-C**. The **topasout** command can be used to generate reports from the data gathered from **topasrec**. With the **topas** CEC analyzer you have a nice tool to visualize the data gathered by **topasrec** and **topasout** using graphs in spreadsheets. More details can be found at:

<https://www.ibm.com/developerworks/wikis/display/WikiPtype/topas+CEC+Analyser>

## 9.1.6 The nmon tool

The **nmon** tool is a free tool available from IBM that presents output in a curses-based monitor that is similar to **topas**. **nmon** displays a wealth of information, as shown in Example 9-17 on page 107, about the configuration and utilization of the system, which includes many of the virtualization features.

You can download the **nmon** tool from the **nmon** Wiki website at:

<http://www.ibm.com/developerworks/wikis/display/WikiPtype/nmon>

A complete manual, FAQ, and forum are also at the Wiki website. As of AIX 6.1, **nmon** is also packaged and delivered with AIX.

**Example 9-17 Sample nmon panel**

```

-nmon12e-----2=Top-Child-processor----Host=is3051-----Refresh=2
secs---09:02.11-
Resources -----
OS has 8 PowerPC_POWER6 (64 bit) processors with 8 processors active SMT=On
processor Speed 3504.0 MHz            SerialNumber=104A1B0 MachineType=IBM,9117-MMA
Logical partition=Dynamic    HMC-LPAR-Number&Name=51,is3051
AIX Version=5.3.8.3 TL08 Kernel=64 bit Multi-Processor
Hardware-Type(NIM)=CHRP=Common H/W Reference Platform Bus-Type=PCI
processor Architecture =PowerPC Implementation=POWER6_in_P6_mode
processor Level 1 Cache is Combined Instruction=65536 bytes & Data=65536 bytes
Level 2 Cache size=not available    Node=is3051
Event= 0 ---      ---      SerialNo 0Id=---      Current=C4A1B0 When=---
processor-Utilisation-Small-View -----Entitledprocessor= 0.40
Usedprocessor= 1.815-----
PURR Stats ["-"=otherLPAR] 0-----25-----50-----75-----100
processor User% Sys% Wait% Idle%|
|
0 56.0 0.3 0.0 0.1|UUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUU----->
1 33.3 0.0 0.0 0.1|UUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUU----->
2 68.2 23.6 0.0 0.0|UUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUssssssssss----->
3 0.0 0.0 0.0 0.1|>-----|
4 0.0 0.0 0.0 0.0|>-----|
5 0.0 0.0 0.0 0.0|>-----|
6 0.0 0.0 0.0 0.0|>-----|
7 0.0 0.0 0.0 0.0|>-----|
EntitleCapacity/Virtualprocessor
+-----+-----+-----+-----+
EC+ 86.6 13.2 0.0 0.1|UUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUssssss|
VP 39.3 6.0 0.0 0.1|UUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUss-----|
EC= 453.7% VP= 45.4% +--No Cap---|--Folded=2--|-----100% VP=4
processor+
Memory -----
Physical PageSpace | pages/sec In Out | FileSystemCache
% Used 42.8% 0.2% | to Paging Space 0.0 0.0 | (numperm) 4.0%
% Free 57.2% 99.8% | to File System 0.0 0.0 | Process 23.2%
MB Used 877.4MB 3.5MB | Page Scans 0.0 | System 15.7%
MB Free 1170.6MB 2044.5MB | Page Cycles 0.0 | Free 57.2%
Total(MB) 2048.0MB 2048.0MB | Page Steals 0.0 | -----
| Page Faults 656.5 | Total 100.0%
-----|
Min/Maxperm 84MB( 4%) 168MB( 8%) <--% of RAM | numclient 4.0%
Min/Maxfree 960 1088 Total Virtual 4.0GB | maxclient 8.2%
Min/Maxpgahead 2 8 Accessed Virtual 0.7GB 18.5% | User 11.8%
Min/Maxpghead 2 8 Accessed Virtual 0.7GB 18.5% | Pinned 26.5%
Top-Processes-(88) -----Mode=3 [1=Basic 2=processor 3=Perf 4=Size 5=I/O
6=Cmnds]--
PID %processor Size Res Res Res Char RAM Paging Command
Used KB Set Text Data I/O Use io other repage
385030 88.8 112 116 4 112 0 0% 0 0 0 spin
442520 60.0 112 116 4 112 0 0% 0 0 0 spin
295062 31.6 1248 1316 340 976 10 0% 0 655 0 xterm
348376 0.1 7212 7536 532 7004 0 1% 0 0 0 nmon12e_aix537

```

You can toggle on or off various sections that contain different metrics. Example 9-18 shows section possibilities that are offered when you depress the h key.

*Example 9-18 nmon section options*

---

h = Help information	q = Quit nmon	0 = reset peak counts
+ = double refresh time	- = half refresh	r =
Resourcesprocessor/HW/MHz/AIX		
c = processor by processor	C=upto 128 processors	p = LPAR Stats (if LPAR)
l = processor avg longer term	k = Kernel Internal	# = Physicalprocessor if SPLPAR
m = Memory & Paging	M = Multiple Page Sizes	P = Paging Space
d = DiskI/O Graphs	D = DiskIO +Service times	o = Disks %Busy Map
a = Disk Adapter	e = ESS vpath stats	V = Volume Group stats
^ = FC Adapter (fcstat)	0 = VIOS SEA (entstat)	v = Verbose=OK/Warn/Danger
n = Network stats	N=NFS stats (NN for v4)	j = JFS Usage stats
A = Async I/O Servers	w = see AIX wait procs	"="= Net/Disk KB<-->MB
b = black&white mode	g = User-Defined-Disk-Groups (see cmdline -g)	
t = Top-Process --->	1=basic 2=processor-Use 3=processor(default) 4=Size	
5=Disk-I/O		
u = Top+cmd arguments	U = Top+WLM Classes	. = only busy disks & procs
W = WLM Section	S = WLM SubClasses)	

---

You can optionally save the nmon data in a file, which you can then give as input to the nmon analyzer, which uses MS Excel to produce dozens of graphs. The nmon analyzer is also available at the nmon Wiki website.

## 9.2 Performance monitoring on IBM i

Performance tools for the IBM i operating system can be divided into four major groups:

- ▶ Ad-hoc performance tools that show a snapshot of hardware resource utilization
- ▶ Data collection tools to collect data about hardware resource utilization continuously
- ▶ Performance analysis tools to format, view, and summarize the collected data
- ▶ Enhanced tools for root cause analysis of performance problems.

This chapter can only provide a quick overview of the most commonly used performance tools on IBM i. For more detailed information about the available tools, see *IBM PowerVM Virtualization Managing and Monitoring*, SG24-7590 and the “IBM i and System i Information Center” at:

<http://publib.boulder.ibm.com/eserver/ibmi.html>.

### 9.2.1 Ad-hoc Performance Tools on IBM i

The WRKSYSSTS command is shipped with the IBM i operating system and can be used to display processor and main storage utilization. When you execute the command with assistance level (ASTLVL) \*INTERMED, you will see a panel as in Figure 9-1 on page 109.

```

Work with System Status
07/12/11 16:13:48 AS0030
% CPU used . . . . . : 361.0 Auxiliary storage:
% DB capability . . . . . : 142.7 System ASP . . . . . : 4233 G
Elapsed time . . . . . : 00:00:01 % system ASP used . . : 64.0045
Jobs in system . . . . . : 5278 Total . . . . . : 4233 G
% perm addresses . . . . . : .224 Current unprotect used : 54733 M
% temp addresses . . . . . : 2.714 Maximum unprotect . . : 66256 M

Type changes (if allowed), press Enter.

System Pool Reserved Max -----DB----- ---Non-DB---
Pool Size (M) Size (M) Active Fault Pages Fault Pages
1 5200.00 470.22 +++++ .0 .0 .0 .0
2 10869.12 37.63 1400 44.9 323.7 425.1 1210
3 291.00 .50 71 .0 .0 .0 .0
4 1459.00 1.28 352 .0 .0 .0 .0
5 1459.00 2.00 356 .0 .0 .0 .0
More...

Command
====>
F3=Exit F4=Prompt F5=Refresh F9=Retrieve F10=Restart F12=Cancel
F19=Extended system status F24=More keys

```

Figure 9-1 Output of WRKSYSSTS

When you use IBM i in a logical partition that is utilizing processor resources from a shared processor pool, you must consider several metrics: The number of physical processors in the shared processor pool, the number of virtual processors assigned to the partition, and the processing units assigned to the partition (also called entitlement). The entitlement indicates the guaranteed processing capacity of the partition.

In Figure 9-1, the shared processor pool has four physical processors, and the partition is configured with three virtual processors and an entitlement of 0.3 processing units. In the WRKSYSSTS panel, the value for “% processor used” indicates the processor utilization in relation to the entitlement of 0.3 processing units. A value of more than 100% is possible because additional processor resources from the shared processor pool can be used. Use function key F19 to get a more detailed view of processor utilization, as in Figure 9-2.

```

Extended System Status
07/12/11 16:13:48 AS0030
Elapsed time . . . . . : 00:00:01
% CPU used . . . . . : 361.0
% DB capability . . . . . : 142.7
% uncapped CPU capacity used . . . . . : 36.1
% shared processor pool used . . . . . : 39.9

F3=Exit F5=Refresh F10=Restart F12=Cancel

```

Figure 9-2 Output using F19

The value for “% uncapped processor capacity used” calculates the current processor utilization in relation to the number of virtual processors (the maximum capacity that this partition can get). The value for “% shared processor pool used” considers how much of the shared processor pool is currently in use by all partitions sharing it, including this one. If this value approaches 100%, the system is becoming processor bound. You can either increase the number of processors in the shared pool or reduce the workload to avoid this situation.

The “Work with System Status” panel also has information about the utilization of main storage pools. The system pools 1 (\*MACHINE pool) and 2 (\*BASE pool) are always configured, while other pools are only shown if you have configured them and assigned them to an active subsystem (see 8.4.1, “Separation of main storage pools” on page 82). You can directly change pool sizes for all main storage pools except the \*BASE pool (system pool 2). To change the pool size, overwrite the current value in column “Pool Size (M)” with the desired value and press Enter. The \*BASE pool gets the storage that is left over from the available main storage after all the other pools have their desired amount of storage assigned. The page fault rates (DB faults per second and non-DB faults per second) are an important indicator for system performance.

The WRKSYSACT command is only available when the licensed program IBM Performance Tools for i (57xx-PT1 - Manager Feature) is installed. A sample panel looks like Figure 9-3.

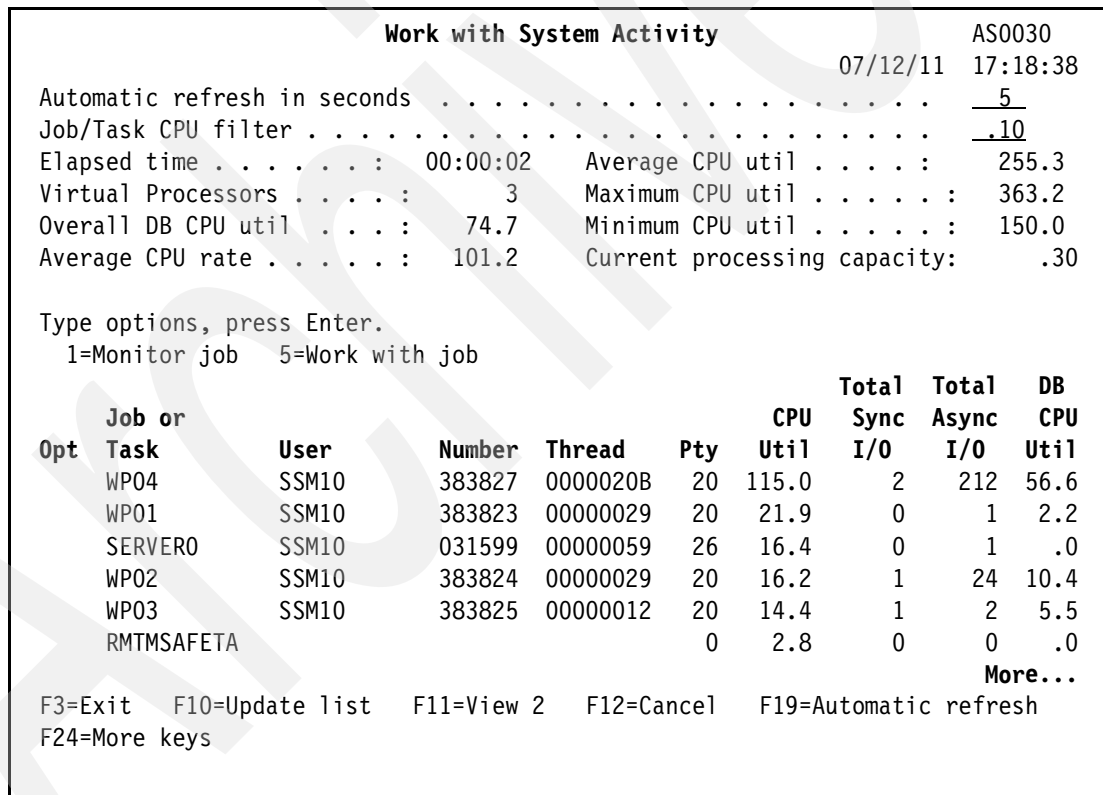


Figure 9-3 Output of WRKSYSACT

The Work with System Activity tool is showing the number of virtual processors and the processing capacity (entitlement), which you could otherwise only see in the Hardware Management Console (HMC). In addition, the tool lists processes and Licensed Internal Code (LIC) tasks that use most of the processor resources.



## 9.2.2 Performance data collection tools on IBM i

Performance data collection on IBM i is performed by the Collection Services, which are shipped as part of the operating system. You can control the Collection Services through the command line interface, application programming interfaces (APIs), the System i Navigator interface and the IBM Systems Director Navigator Performance interface (available with IBM i 6.1 or later).

In the command line interface, use the STRPFCOL command with parameter COLPRF to start a performance collection based on a collection profile. The collection profile defines what type of performance data to collect. Multiple collection profiles are available. When running SAP applications in a virtualized environment, the suggested collection profile is \*STANDARDP. The values \*MINIMUM and \*STANDARD are not recommended because they do not include LPAR and communication data. You can change other attributes of the performance data collection, such as the default collection interval, the cycle time and interval, and the collection retention time, by using the CFGPFCOL command. If you want to end a performance data collection, use the ENDPFCOL command.

The collected data is initially stored in so-called management collection objects of type \*MGTCOL. They are stored in a library that you can define with the CFGPFCOL command, typically library QPFRDATA or library QMPGDATA. After some time, typically every day at midnight, the collection is cycled, that is, the Collection Services create a new management collection object and write further performance data into the new object. You can now convert the old management collection object into database files for further analysis with the CRTPFRTA command. After that, you can use the CRTPFRSUM command to create additional indexes on the performance database files and to allow faster access to the summary data in the performance analysis tools. In order to automate the creation of these database files, use the CFGPFCOL command and set both the parameters “Create database files” (CRTDBF) and “Create performance summary” (CRTPFRSUM) to \*YES.

When the licensed program “IBM Performance Tools for i” is installed, you can use option 2 of menu PERFORM (“Collect performance data”) to start, end, or configure the performance data collection.

With the System i Navigator interface, you can start, end, configure, and manage the Collection Services through the path Configuration and Service → Collection Services as shown in Figure 9-4 on page 112.

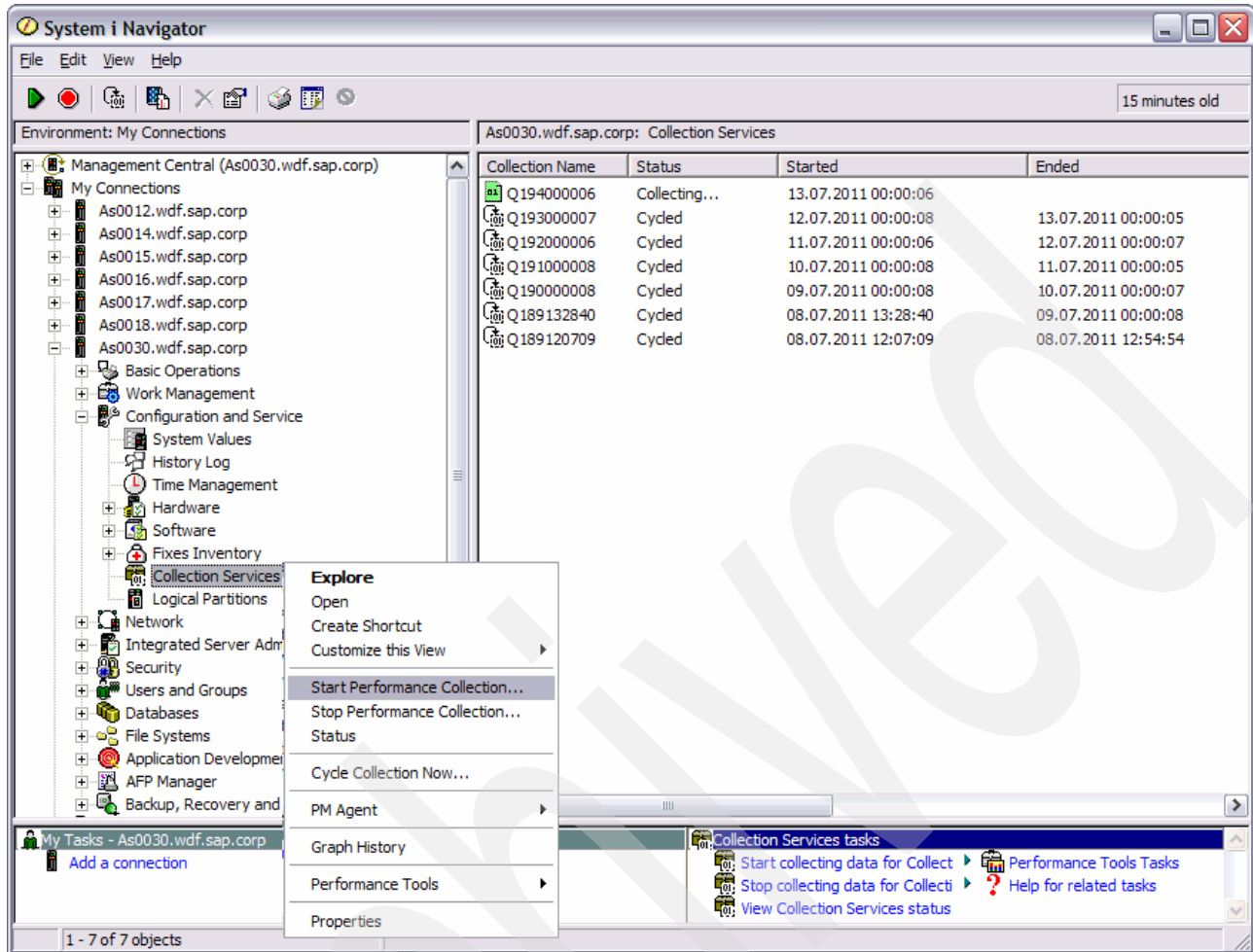


Figure 9-4 System i Navigator for Collection Services

The Collection Services can be started with one of the shipped collection profiles, but it is also possible to individually select what data to collect in which intervals, as shown in Figure 9-5 on page 113.

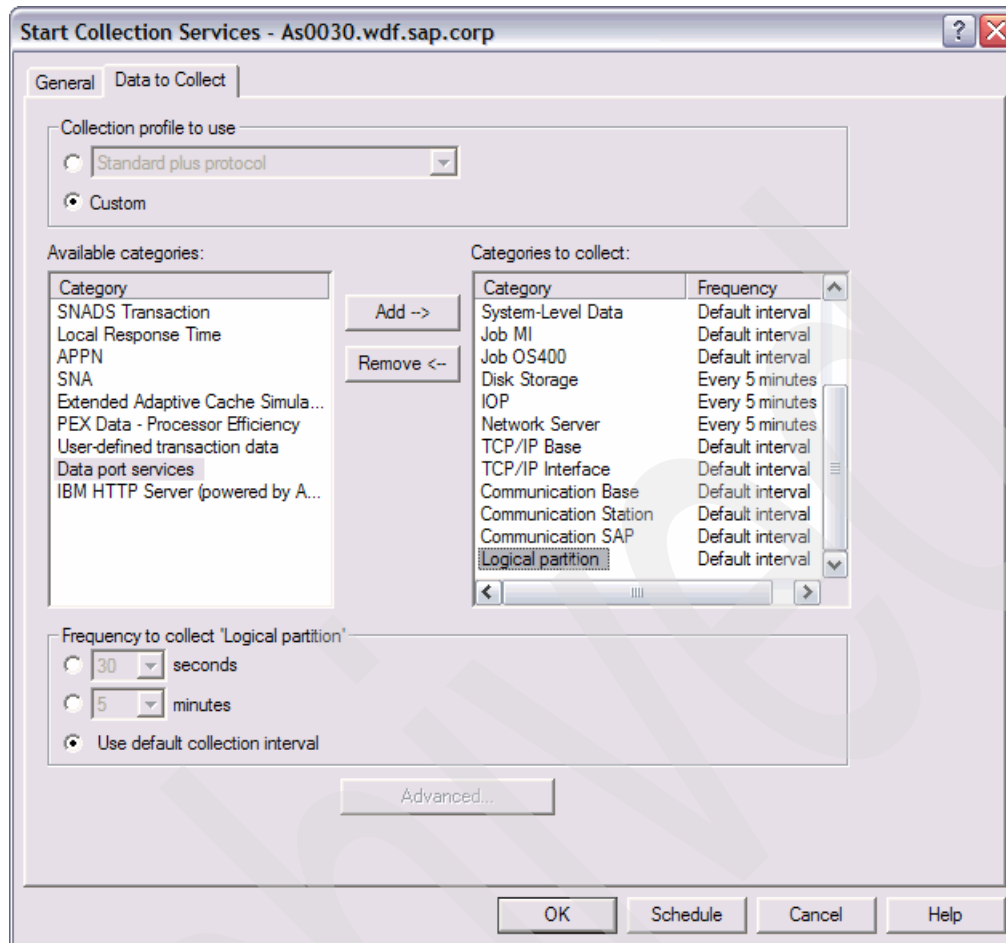


Figure 9-5 Selection panel for Collection Services options

### 9.2.3 Performance Data Analysis Tools on IBM i

The IBM i operating system offers three ways to analyze the data that was collected by the Collection Services: Querying the output tables of the CRTPFRTA command, using the Graph History function of the System i Navigator, or using the “Performance Data Investigator” of the IBM Systems Director Navigator for i Performance (available with IBM i 6.1 or later). You can perform a much more detailed analysis of the collected performance data with the licensed program IBM Performance Tools for i (57xx-PT1 - Manager Feature).

The formats and columns of the Collection Services output tables are documented in the IBM i and System i Information Center at:

<http://publib.boulder.ibm.com/eserver/ibmi.html>

Follow the path under **Systems management** → **Performance** → **Reference information for Performance** → **Collection Services data files**. For SAP applications on IBM i in a virtualized environment, tables QAPMLPARH and QAPMSHRMP are of special interest. Table QAPMLPARH contains logical partition configuration and utilization data and is available in IBM i 6.1 and higher. Table QAPMSHRMP reports shared memory pool data and is available in IBM i 7.1 and higher when a partition is defined to use a shared memory pool. This is supported on POWER6 or later with firmware level xx340\_075 or later. Table QAPMLPAR contains cross-partition data, which is collected by the IBM Systems Director. While the IBM Systems Director Server is running on another server or partition, the platform

agent support for IBM i is provided with licensed program IBM Universal Manageability Enablement for i (57xx-UME).

You can display performance data in the System i Navigator through the path **Configuration and Service** → **Collection Services**. Select a collection and then choose the option **Graph History**. A variety of metrics can be selected for display, for example the average processor utilization, as shown in Figure 9-6. Note that the processor utilization is given in percent of the entitlement, so values larger than 100% are possible when running in a virtualized environment with a shared processor pool.

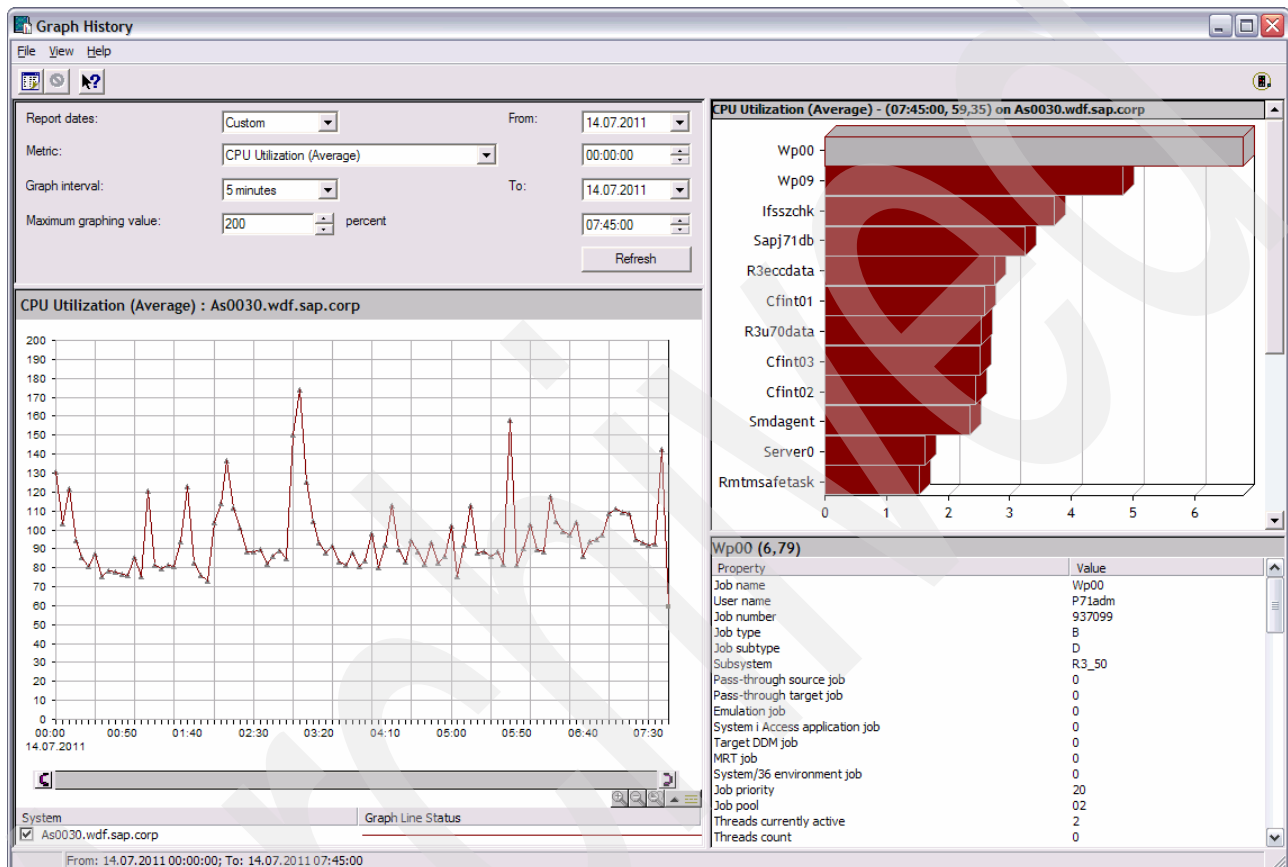


Figure 9-6 Performance Graph History for processor utilization

You can execute the Performance Data Investigator by following these steps:

1. Access this URL from a web browser:  
`http://<server name>:2001`
2. Log in with a user ID and password.
3. Select **Performance** from the left panel.
4. Select **Investigate Data**.
5. Select **Health Indicators**.
6. Select a performance collection and click **Display**.

You will get output similar to that shown in Figure 9-7 on page 115. You can change thresholds and select a variety of views with the Select Action button.

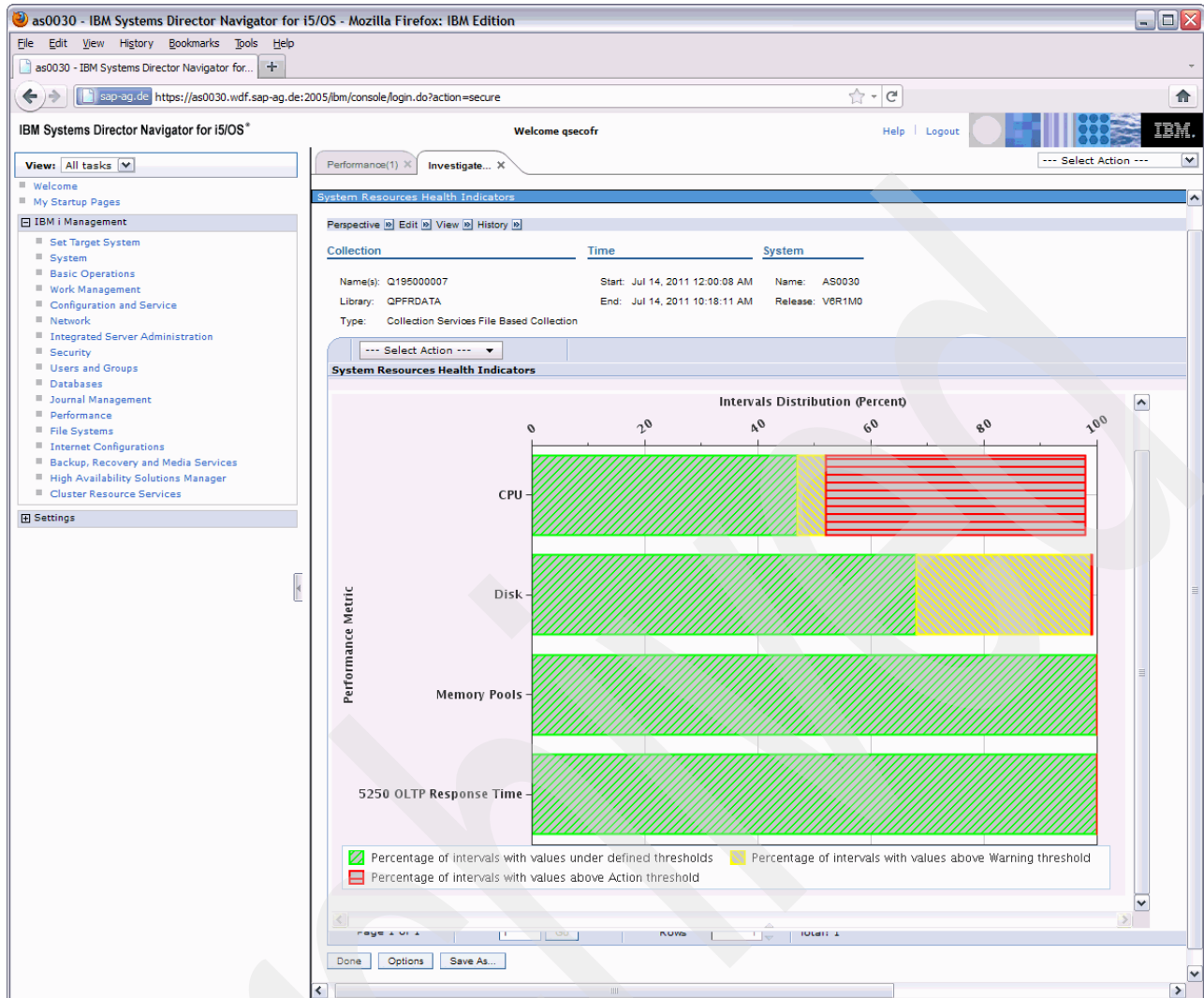


Figure 9-7 Performance Data Investigator in the Systems Director Navigator

The licensed program IBM Performance Tools for i offers a variety of reports and analysis tools for a more convenient analysis of the performance data collected by the Collection Services. You can access these tools through the menu PERFORM or the STRPFRT command.

When using the STRPFRT command, you can select the library that contains the performance data, if it is different from QPFRDATA. In the resulting menu, you can use option 3 (“Print performance report”) to print system reports for a general overview or component reports for a closer look at individual hardware resources. In a virtualized environment, these reports show the number of virtual processors and the processor units (entitlement). Processor utilization is reported in percent of the entitlement, so values above 100% are possible. The system report and the component report for disk activity include the disk service time, which can be an important indicator of a potential bottleneck when using external storage or a Virtual I/O Server.

Option 7 (“Display performance data”) in the Performance Tools menu enables you to navigate through a complete data collection or a certain interval within the data collection. After selecting a data collection and an interval, you will see the overview panel shown in Figure 9-8 on page 116.

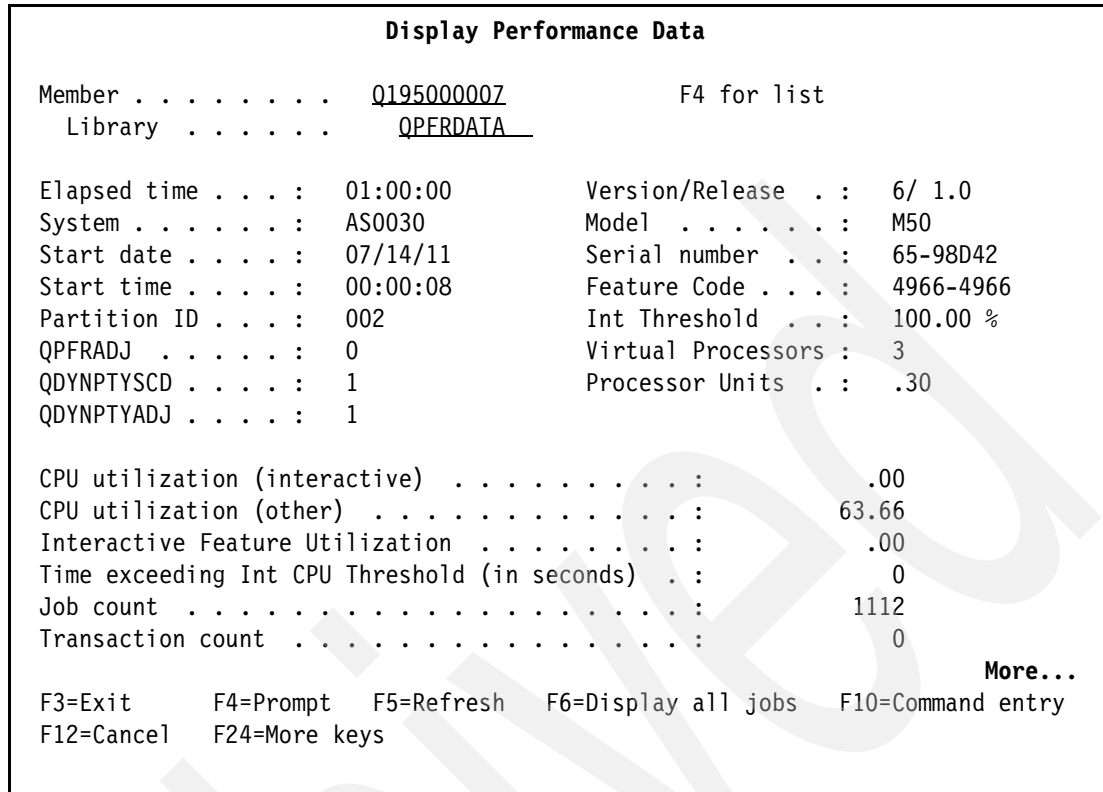


Figure 9-8 Display Performance Data

From this panel, you can group the data by subsystem, job type, interval, main storage pool, disk unit, or communication line, or you can display a list of all processes that were active during the collection period. From that list you can see hardware resource consumption and wait times for each process individually.

## 9.2.4 Enhanced tools for root cause analysis

The previously mentioned tools are designed to collect data continuously, but they can only display summary information. This may be sufficient to identify, for example, lack of processor resources as the main cause of poor performance, but it does not help you to identify the software components that are using most of this resource. For a more in-depth analysis, you can use the following tools:

- ▶ Performance Explorer (PEX)
- ▶ IBM i Job Watcher
- ▶ IBM i Disk Watcher

### **Performance Explorer (PEX)**

The Performance Explorer allows very specific and comprehensive data collections in the shape of statistics or traces, or based on profile information. The data collection can be system wide or process specific. Usually it produces a considerable amount of output, so it should only be active for a short amount of time. You define the type and amount of data collected with the ADDPEXDFN command. You start and stop the actual data collection with the commands STRPEX and ENDPEX. You can evaluate the data with the PRTPEXRPT command or with the tool "IBM iDoctor for IBM i". Beginning with IBM i 6.1, all components except for "IBM iDoctor for IBM i" are included in the base operating system.

### ***IBM i Job Watcher***

The IBM i Job Watcher can be used to collect call stacks, SQL statements, wait statistics and objects being waited on for all processes on the system, based on periodic snapshots. The data collection can be system wide or process specific and should only be active for a short amount of time. You define the type of data collected with the ADDJWDFN command. You start and stop the actual data collection with the commands STRJW and ENDJW. You can evaluate the data with the tool “IBM iDoctor for IBM i”. Beginning with IBM i 6.1, the commands ADDJWDFN, STRJW, and ENDJW are shipped with the base operating system.

### ***IBM i Disk Watcher***

The IBM i Disk Watcher can be used to collect data for I/O operations to disk units along with data that makes it possible to identify which processes have caused them and which objects were related. The data collection is system wide, but you can limit it to a specific auxiliary storage pool (ASP). You define the type of data collected with the ADDDWDFN command. You start and stop the actual data collection with the commands STRDW and ENDDW. You can evaluate the data with the tool “IBM iDoctor for IBM i”. Beginning with IBM i 6.1, the commands ADDDWDFN, STRDW, and ENDDW are shipped with the base operating system.

More information about these tools can be found in the IBM i and System i Information Center at:

<http://publib.boulder.ibm.com/eserver/ibmi.html>

under **Systems management** → **Performance** → **Applications for performance management** → **Performance data collectors**. Information about the IBM iDoctor for IBM i can be found at:

[http://www-912.ibm.com/i\\_dir/idoctor.nsf/iDoctor.html](http://www-912.ibm.com/i_dir/idoctor.nsf/iDoctor.html).

This page also contains links to *Application and Program Performance Analysis Using PEX Statistics on IBM i5/OS*, SG24-7457 and to *IBM iDoctor for iSeries® Job Watcher: Advanced Performance Tool*, SG24-6474.

## **9.3 Other monitoring tools**

In this section, we cover other monitoring tools, such as LPAR2rrd and Ganglia.

### **9.3.1 LPAR2rrd**

To monitor a virtualized environment, LPAR2rrd provides a good overview of used processor and memory resources across LPARs with minimal overhead. The advantage of this tool is that it does not require that you install agents into the LPARs on the system. Instead it evaluates, accumulates, and presents the utilization data from the HMC. It is independent from the operating systems running inside the monitored LPARs. The tool itself has some restrictions on where it can run and these are documented here. Because the HMC does not gather utilization data from inside the LPARs (intentionally), this tool is perfectly suited for environments that use mostly shared pool LPARs. After you set up the tool, the only activity to add new physical machines to the monitoring environment is to enable gathering of utilization data for the server on the HMC.

The features of this tool are:

- ▶ It creates charts based on utilization data collected on HMCs.
- ▶ It is intended mainly for micropartitioned systems.

- ▶ It does not require to install clients into LPARs.
- ▶ It is independent from the OS installed in the LPARs.
- ▶ It is easy to install and configure.
- ▶ It can also deal with LPARs moved with Live Partition Mobility (for example, Ganglia has some issues with this case).

In addition to these advantages, here are more facts:

- ▶ It will not provide processor utilization data for dedicated LPARs.
- ▶ For Dedicated Shared LPARs, it shows only how many cycles are donated to the shared pool. Because these are typically the idle cycles of an LPAR, you get a pretty good impression of the processor usage in a Dedicated Shared LPAR (in contrast to the dedicated LPAR).

The representation of the gathered data occurs through a small web application that provides a simple navigation through all machines and LPARs in your environment and presents the resource usage data as graphs. Figure 9-9 shows one example of a machine using mainly LPARs in the shared pool. It shows the accumulated processor usage of all LPARs in the shared pool.

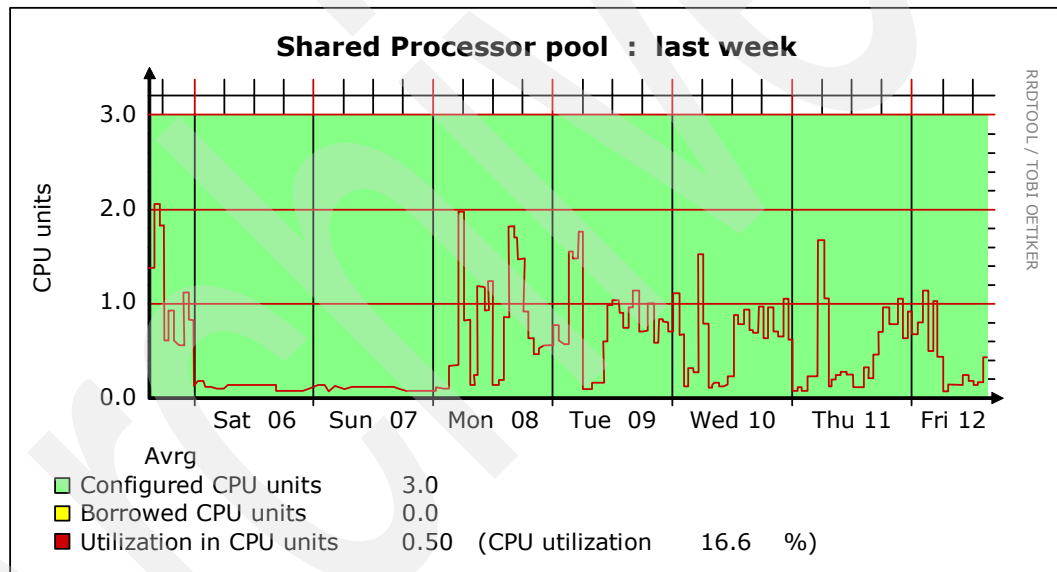


Figure 9-9 Web front end display of LPAR2rrd

The tool provides similar graphs also for each single LPAR and for different time windows (day, week, month, and year). Current information about the latest available version, more images, an explanation of the output, and detailed installation and set-up descriptions are provided at:

<http://www.ibm.com/developerworks/wikis/display/virtualization/LPAR2rrd+tool>

LPAR2rrd requires a UNIX system with the following prerequisites:

- ▶ Web server (for example, Apache)
- ▶ Perl
- ▶ ssh
- ▶ RRDTTool



LPAR2rrd was tested with AIX 5.3/5.2, SuSE Linux, RedHat and Ubuntu on Intel. It is not necessary to dedicate the UNIX system exclusively for the tool.

To retrieve the LPAR utilization data from the HMC, the HMC version that you are using must be greater or equal to V5R2.1, and the firmware on the managed system must be at least on SF240\_201.

You can also use the web front end of the HMC to enable the collection of utilization data. Our original procedure describes how to do it on the command line in the HMC.

Figure 9-10 through Figure 9-14 on page 120 show the procedure of collection of utilization data on an HMC with the version V7R7.3:

1. In System Management, select the tasks of a specific server, as shown in Figure 9-10.

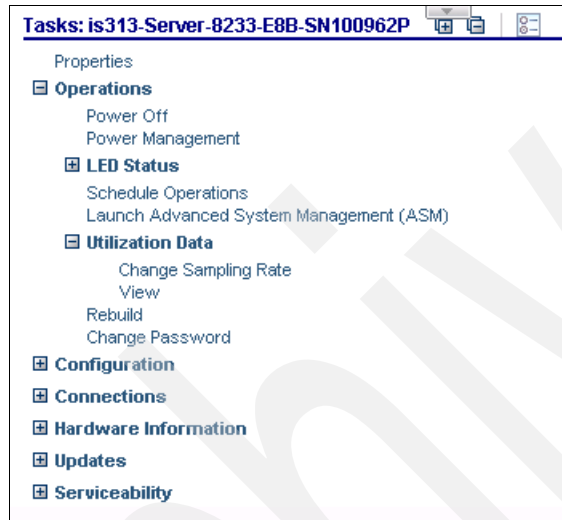


Figure 9-10 HMC procedure

2. Select **Operations** → **Utilization Data**, and choose the option **Change Sampling Rate**, as shown in Figure 9-11. When you choose this option, you can select the sampling frequency for your data.

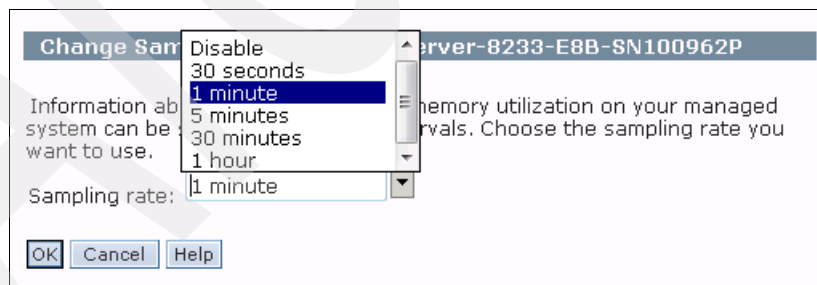


Figure 9-11 HMC - select sampling rate

3. Alternately, instead of choosing the option **Change Sampling Rate**, you can choose the option **View** as shown in Figure 9-12 on page 120. Using this option, you can verify that the HMC is gathering the data as expected. After choosing the **View** option you can filter the events that you want to see.

Figure 9-12 HMC - filter events

A list of events is displayed, as shown in Figure 9-13, that matches your criteria. You can also show the detail information that is available for each event.

Select	Date	Time	Event
<input type="radio"/>	7/14/11	9:01:07 AM CEST	Utilization sample
<input type="radio"/>	7/14/11	9:00:03 AM CEST	Utilization sample
<input type="radio"/>	7/14/11	8:59:32 AM CEST	Utilization sample
<input checked="" type="radio"/>	7/14/11	8:58:28 AM CEST	Utilization sample
<input type="radio"/>	7/14/11	8:57:23 AM CEST	Utilization sample
<input type="radio"/>	7/14/11	8:56:18 AM CEST	Utilization sample
<input type="radio"/>	7/14/11	8:55:09 AM CEST	Utilization sample
<input type="radio"/>	7/14/11	8:54:04 AM CEST	Utilization sample
<input type="radio"/>	7/14/11	8:53:00 AM CEST	Utilization sample
<input type="radio"/>	7/14/11	8:51:55 AM CEST	Utilization sample

Figure 9-13 HMC - select events

As shown in Figure 9-14, we look at the system data, which is relatively static.

Figure 9-14 HMC - select system

You can also look at other panels such as the LPAR and pool information. Those panels provide very detailed information. Figure 9-15 shows an example of an overview of the utilization of all partitions.

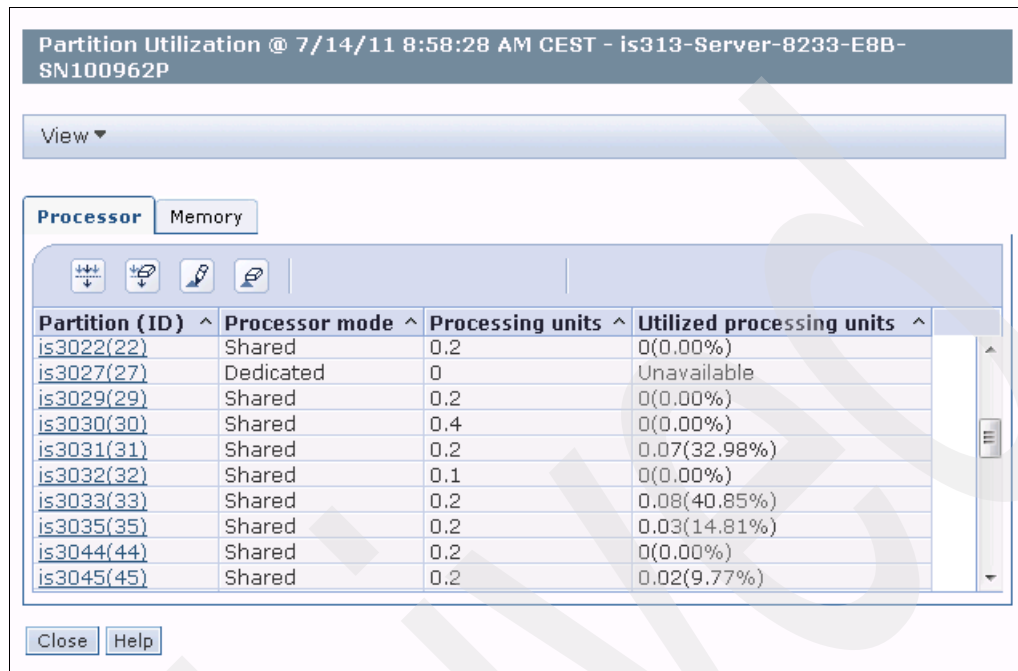


Figure 9-15 HMC: Partition Utilization

In this book, we do not describe the installation procedure for LPAR2rrd itself because the instructions are provided in a simple and straightforward manner on the website at:

<http://www.ibm.com/developerworks/wikis/display/virtualization/LPAR2rrd+tool>

### 9.3.2 Ganglia

Ganglia is an open source distributed monitoring system that was originally targeted at High Performance Computing (HPC) clusters. Ganglia relies on a hierarchical design to scale well, especially in large environments. The freely available code runs on many different operating systems and hardware platforms. Ganglia is easy to set up and can be used through a web front end with any web browser. Unlike most open source monitoring tools, Ganglia is also aware of the virtualized Power Systems environment. More and more clients use Ganglia to monitor their IT landscape with both SAP and non-SAP systems.

#### Architecture overview

The Ganglia monitoring system collects and aggregates performance data from different computing systems. Ganglia defines three terms that describe the level of detail of collected data in a hierarchical structure:

- ▶ Node
- ▶ Cluster
- ▶ Grid

A node is the smallest entity where the actual performance metrics get collected. Usually a node is a small physical system that does a certain task or specific calculation. One or more nodes doing the same task participate in a cluster, and several different clusters make up a

grid. Concerning time information, Ganglia aggregates per default performance data in the following resolutions: last hour, last day, last week, last month, and last year.

In different environments than HPC, you can easily translate Ganglia terms to corresponding terms used therein. In a virtualized environment, you can see a node as an equivalent to a single operating system instance, regardless of the physical occurrence of the node itself. Therefore, a node can be a virtual machine, or in Power Systems terms, a logical partition (LPAR). A cluster is then a group of nodes, or in our case, LPARs that are running on the same physical system. A grid is just a group of physical systems.

Table 9-1 shows the translation of Ganglia terms to other environments.

Table 9-1 Ganglia terms<sup>1</sup>

Ganglia terms	HPC environment	Data Center	Virtualized environment
Node	Physical system	Physical system	Virtual machine (VM) or LPAR
Cluster	Systems with the same task	Grouping of platforms or system type (for example, Web services or all UNIX systems)	All VMs or LPARs on the same physical system (CEC)
Grid	Group of clusters	All clusters in data center	Multiple CECs

The Ganglia monitoring system consists of four main components:

- ▶ Ganglia monitoring daemon
- ▶ Ganglia meta daemon
- ▶ Data store
- ▶ Web front end

Figure 9-16 on page 123 shows the relationship between the four components.

<sup>1</sup> Ganglia, reprinted by permission.

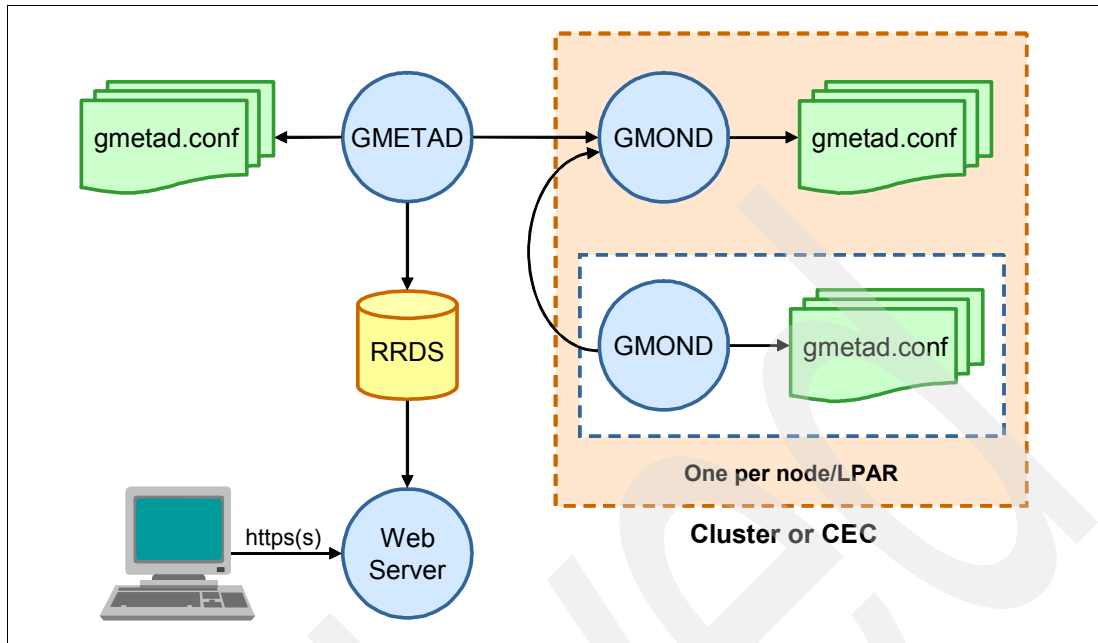


Figure 9-16 Ganglia architecture overview<sup>2</sup>

The Ganglia monitoring daemon (gmond) runs on every node to be monitored. It is a single binary with a single configuration file, `/etc/gmond.conf`.

According to the configuration, the monitoring daemon collects all supported performance metrics on the node repeatedly in certain intervals of time. All monitoring daemons within a cluster send the collected data on to one designated monitoring daemon within the same cluster.

The Ganglia meta daemon (gmetad) collects all of the data from clusters and stores it into data files. The meta daemon has just one single configuration file, `/etc/gmetad.conf`.

This configuration file lists the available data sources. These data sources are usually monitoring daemons that represent one cluster each. In more complex environments, you can have cascaded meta daemons forming a hierarchy according to the three mentioned entity levels.

The data files are in the round robin database (RRD) file format. RRD is a common format for storing time series data and is used by many open source projects. The RRDtool Web site provides you with a very good description of using the RRD file format for generating customized graphs at:

<http://oss.oetiker.ch/rrdtool/>

The web front end runs on a PHP-enabled web server and needs access to the RRD data files of the meta daemon. In most cases, you run the meta daemon and the web server on the same machine. The web front end generates performance graphs on the grid, cluster, and node level in the specified time frame. Hence, the web front end is based on the PHP scripting language; therefore, you can easily customize the interface and the shown graphs.

Through a separate executable called gmetric, you can insert additional performance metrics that are unknown to the Ganglia monitoring daemon.

<sup>2</sup> Ganglia, reprinted by permission.

## Ganglia and Power Systems virtualization

Although Ganglia already runs on a variety of platforms, the current AIX and Linux implementation of Ganglia lacks the support for IBM Power Systems processor-specific metrics. To fully support the Power Systems platform, the relevant portion of Ganglia source code was adapted by Nigel Griffith and Michael Perzl. They also provided a proposal to include the Power metrics into future Ganglia releases.

Table 9-2 shows the metrics that the AIX and Linux version of Ganglia supports<sup>3</sup>.

Table 9-2 Metrics of AIX and Linux supported by Ganglia

Ganglia metric terms	Value type	Description
Capped	boolean	Specifies if the LPAR is capped (0 = false, 1=true)
processor_entitlement	decimal number	Specifies the number of processor units (for example 1.5)
processor_in_LPAR	integer	Number of virtual processors for a micropartition, number physical processors for a dedicated LPAR, or standalone system
processor_in_machine	integer	Number of physical processors in the system
processor_in_pool	integer	Number of physical processors in the shared processor pool for micropartitions
processor_in_idle	decimal	Number of processor units that are idle in the shared processor pool
processor_used	decimal	Number of processor units utilized by the LPAR
disk_read	integer	Number of KBs read from disk
disk_write	integer	Number of KBs written to disk
kernal 64-bit	boolean	Specifies if the LPAR is capped (0 = false, 1=true)
Lpar	boolean	Specifies if the LPAR is capped (0 = false, 1=true)
LPAR_name	string	Specifies the name of the LPAR given in the HMC
LPAR_num	integer	Specifies the LPAR number
Oslevel	string	Specifies the level of operating system
serial_num	string	Specifies the serial number of the physical system
Smt	boolean	Specifies if simultaneous multithreading is turned on (0 = false, 1=true)
SpLPAR	boolean	Specifies if the LPAR is capped (0 = false, 1=true)
Weight	integer	Specifies the weight of the LPAR

**Note:** For monitoring the processor\_pool\_idle metric in an LPAR, grant the pool utilization authority in the Hardware Management Console.

<sup>3</sup> Ganglia, reprinted by permission

## Downloading and installing Ganglia

The Ganglia source code and precompiled binaries for different platforms are available at the Ganglia project home page:

<http://ganglia.info/>

You can download precompiled binary packages for AIX and Linux under Power Systems with the POWER Ganglia extensions from:

<http://www.perzl.org/ganglia>

To install Ganglia, follow the instructions on the Ganglia project home page.

## Sample usage

Figure 9-17 shows a usage example of Ganglia. Three LPARs (Is3772, Is3773, and Is3752) running on the same managed system are grouped as a Ganglia cluster (LoP 7.4).

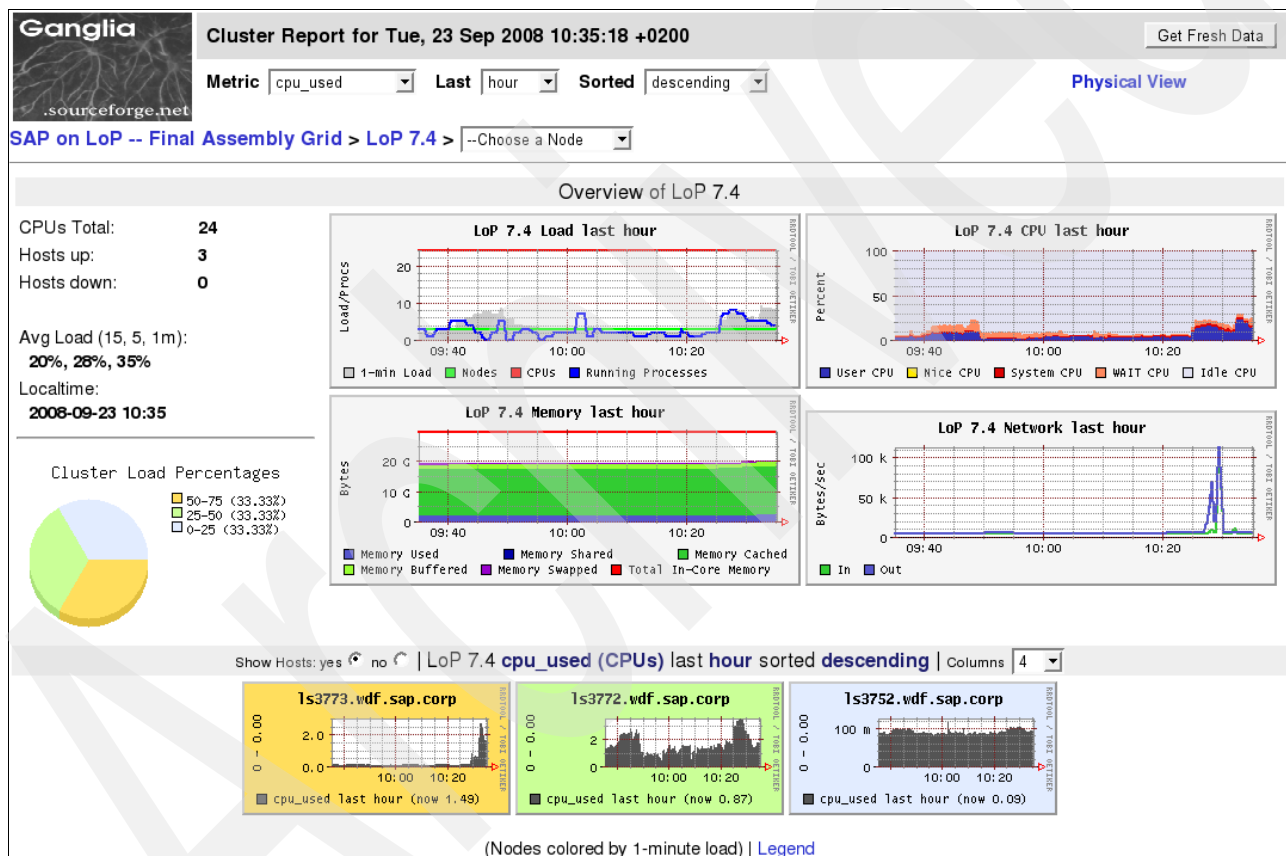


Figure 9-17 Ganglia usage example<sup>4</sup>

You can access a live demo of Ganglia from:

<http://monitor.millennium.berkeley.edu/>

<sup>4</sup> Ganglia, reprinted by permission.

### 9.3.3 Tivoli Monitoring

Besides the introduced monitoring tools, IBM provides a mighty monitoring facility within its Tivoli® product family. You can get further details from the deployment guide *IBM Tivoli Monitoring V6.2*, SG24-7444 at:

<http://www.redbooks.ibm.com/abstracts/SG247444.html>

Figure 9-18 shows an example, using the CEC agent, within Tivoli Monitoring, which displays the entire processor and memory allocations per logical partition (LPAR) on a p595.

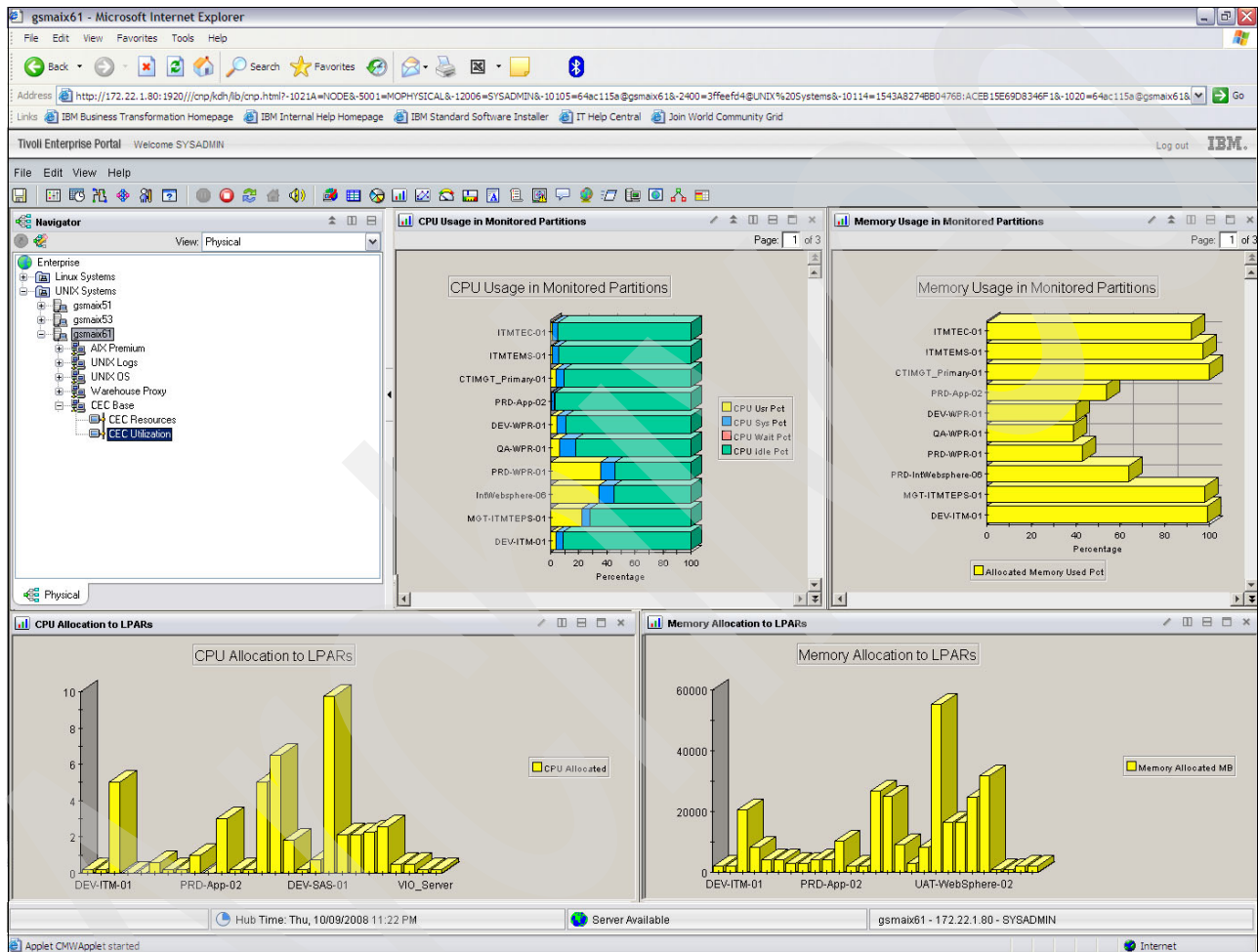


Figure 9-18 Tivoli Enterprise Portal

It is possible to view which LPARs are using the most memory and processor capacity. The workspace also shows how much processor capacity and memory is allocated to each LPAR on the p595 in the same view. Tivoli Monitoring manages your IT infrastructure, which includes operating systems, databases, and servers across distributed and host environments through a single customizable workspace portal.



## 9.4 Monitoring of virtualized SAP systems

In this section, we discuss:

- ▶ Motivation and challenges
- ▶ CCMS transaction st06
- ▶ SAP monitoring infrastructure

### 9.4.1 Motivation and challenges

With the availability of micropartitions and the shared processor pool, SAP systems can no longer rely on the traditional utilization measurements to depict the capacity consumption of the system. The SAP monitoring component (CCMS) depended on these traditional metrics, which were common to most derivatives of UNIX, to provide a common monitoring view of general system resources. In the virtualized landscape, these metrics are still available, but one can no longer rely on them in isolation to come to processor performance conclusions.

On AIX and Linux systems, the SAP operating system data collector, `saposcol`, presents the logical processors as processors. In a dedicated LPAR, for example, if simultaneous multithreading is active, the number of processors increases the physical processors by a factor of the number of threads. In an SPLPAR, the number of processors is a factor of the SMT threads and the online number of virtual processors. In CCMS, it was not possible to differentiate between a non-SMT system with four processors for example, an LPAR with two processors and SMT, or an SPLPAR with two virtual processors and SMT.

In SPLPARs, the processor utilization metrics change their point of reference as opposed to that of a dedicated server. The processor utilization for an uncapped micropartition is relative to its entitlement until it reaches 100% of entitlement and then it remains at 100%. If an SPLPAR is using many times its entitlement, this is not visible within CCMS. Therefore, response times can vary considerably on an SAP system, although the processor utilization remained the same: 100% utilization. This can happen if the SPLPAR got many times its entitlement sometimes, and only its basic entitlement at other times, due to oscillation in contention with other active SPLPAR workloads. This situation makes problem determination an extreme challenge.

The recording of history data for processor utilization, which also only targeted the traditional metrics, was of little value in determining actual use and trends. It was not even possible to determine from this data whether the server was dedicated, and the metrics might therefore still have some reliability, or running in virtualization, with an entirely different point-of-reference.

Virtualization brings many challenges to monitoring and these are further complicated by landscape trends within the SAP infrastructure itself. SAP CCMS collects history data based on aggregation of the hourly and daily system resource utilization. This data is used for trend analysis, regression testing, and performance monitoring. This data is recorded per application server and based on the SAP system, not the server. With the SAP “Adaptive Computing” initiative, you can move SAP components around the infrastructure landscape according to requirement, even landing on servers with different processor types: POWER4 to POWER6 for example.

Using dynamic LPAR functions, you can change the number of processors on the fly, so the number of configured processors can vary on the same dedicated server. For shared processor pools, the number of processors in the pool can change dynamically: dedicated LPARs might release processors that then join the shared pool, on-demand resources can be

brought online, static LPARs might be started and thereby take processors out of the pool, and processors can be dynamically allocated to static LPARs and leave the pool.

The traditional SAP CCMS did not gather any history data on the number of processors, the processor type, or processor speed, related to the utilization data being gathered. Basically we can possibly say that it used 75% of capacity, but we cannot say what the capacity was neither in number nor in speed.

History data coming from CCMS is used by capacity planning tools, accounting methods, and for problem determination support. When this data is no longer relevant, or is even misleading, there are far reaching consequences.

The new integration of virtualization views into CCMS addresses these issues and re-establishes order in what threatened to become chaos. The following section describes the approach to take to integrate virtualization view into CCMS and related tools.

### **9.4.2 SAP CCMS operating system monitor**

The new SAP operating system monitor provides additional detailed configuration and metric information for virtualization and other technologies that could not be viewed previously. In addition, the layout has been redesigned for improved accessibility, as seen in Figure 9-19 on page 129.

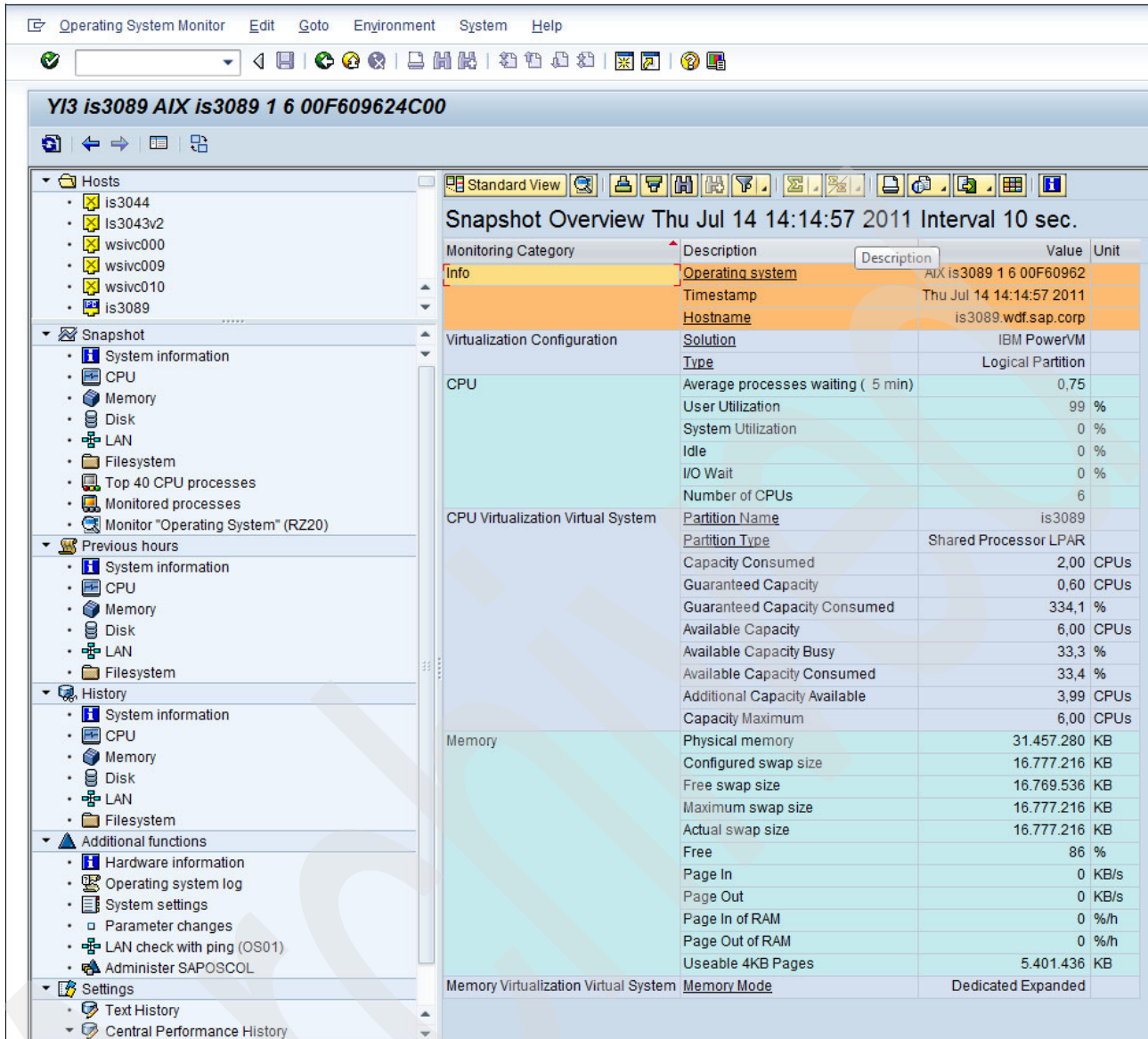


Figure 9-19 Example of st06, standard view<sup>5</sup>

Depending on the SAP Basis release and the support packages installed, the new SAP operating system monitor might be called ST06, OS07, or OS07n. For the remainder of this section, we refer to the new operating system monitor as ST06.

**Tip:** Refer to the following SAP Notes to determine the appropriate transaction call.  
 1084019: st06: New operating system monitor  
 1483047: "Old" version of operating system monitor not available

In Figure 9-20 on page 130, it can be seen that in addition to the new tabular layout, there are various subsections such as "processor", processor Virtualization Virtual System", "Memory", and so on, each containing a set of configuration values and metrics. In the standard view, the sections and number of values and metrics are limited to the minimum, which should help to identify whether problems currently exist in the running system. The expert view, which

<sup>5</sup> Copyright SAP AG 2011. All rights reserved.

provides the complete metric view, can then be used to resolve the issue. Figure 9-20 shows the expert view.

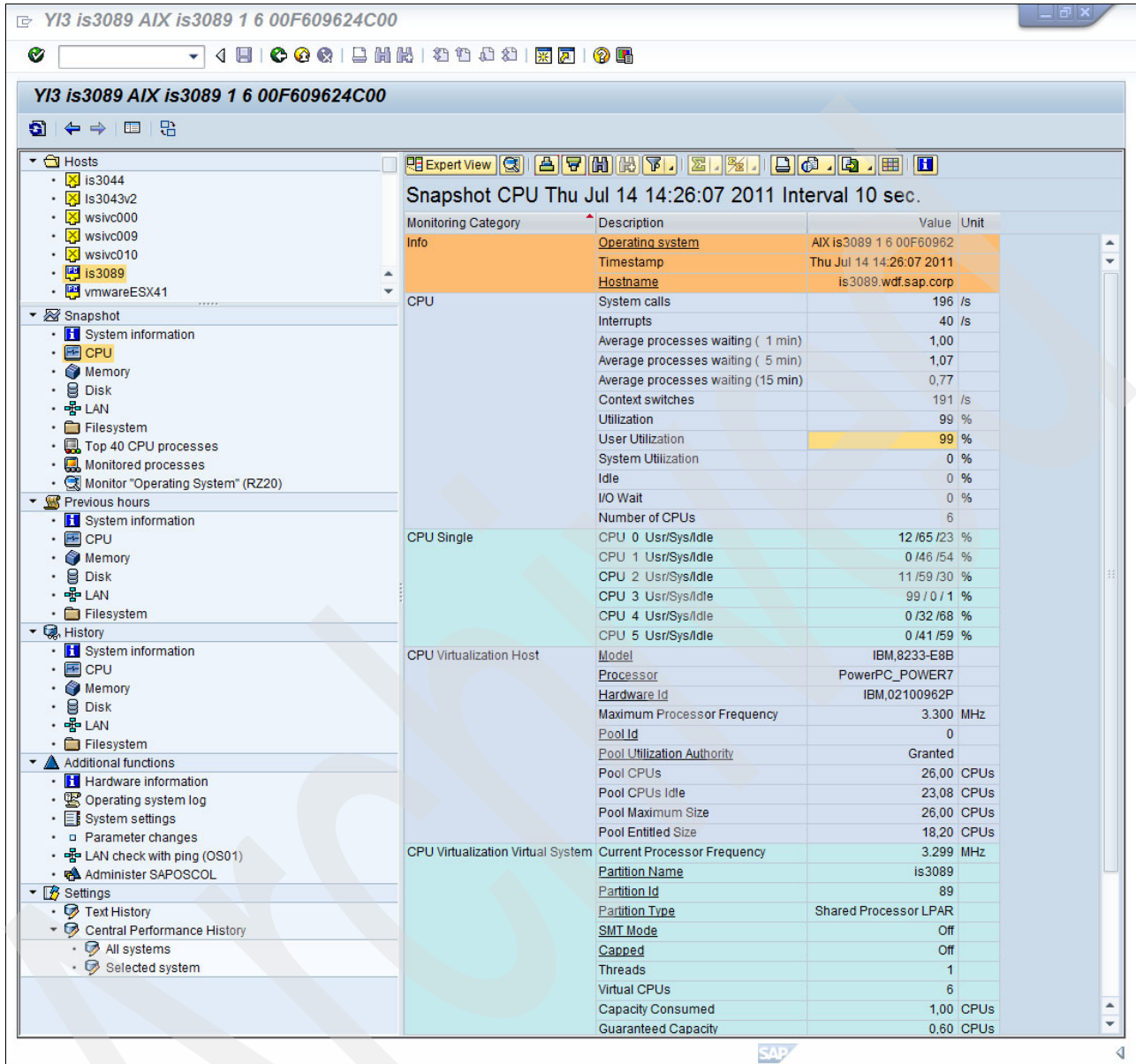


Figure 9-20 Example of ST06, expert view<sup>6</sup>

Most of the new values and metrics that can be seen in the new st06 are defined and delivered by the SAP operating system collector, saposcol. This makes it possible to provide new metrics to st06 without the need to apply new Basis Support Packages, but instead simply by updating saposcol. So, as new PowerVM and operating system technologies become available, it is recommended to keep up-to-date with the latest saposcol version, which usually can be found in the most recent kernel releases. Note that saposcol is downward compatible with respect to the SAP kernel release.

<sup>6</sup> Copyright SAP AG 2011. All rights reserved.

**Tip:** Refer to the following SAP Note on where to download the latest appropriate version of saposcol for AIX.  
710975: FAQ: which saposcol should be used on AIX

The saposcol version information consists of two parts: a platform-independent version string and a platform-dependent version string.

For example,

```
$ saposcol -v | grep version  
SAPOSCOL version COLL 21.03 720 – AIX v12.51 5L-64 bit 110419, 64 bit, single  
threaded, Non-Unicode
```

In this case, the platform-independent version is 21.03, and the platform-dependent version is v12.51.

As of version v12.51, the following new subsections, configuration values, and metrics are available:

The subsection processor Virtualization Host presents processor information relevant to and/or defined on the host.

- ▶ **Model:** Host server model type.
- ▶ **Processor:** Host server processor type.
- ▶ **Maximum Processor Frequency:** Host server nominal processor speed.
- ▶ **Pool Id:** The shared processor pool number for the LPAR
- ▶ **Pool Utilization Authority:** This field indicates whether the LPAR has the authority to retrieve information about the shared pool, for example, the idle processor capacity in the shared pool. Possible values are:
  - Granted
  - Not granted
- ▶ **Pool processors:** This field shows the total number of physical processors in the shared pool to which this LPAR belongs.
- ▶ **Pool processors Idle:** This field indicates the idle capacity of the shared pool in units of physical processors. This value is only available if pool utilization authority is granted.
- ▶ **Pool Maximum Size:** A whole number representing the maximum number of physical processors this pool can contain.
- ▶ **Pool Entitled Size:** The guaranteed size of the processor pool.

Subsection processor Virtualization Virtual System presents processor information relevant to and/or defined on the LPAR.

- ▶ **Current Processor Frequency:** The current operating processor frequency of the LPAR.
- ▶ **Partition Name:** The HMC-defined LPAR name.
- ▶ **Partition Id:** The HMC-defined LPAR number.
- ▶ **Partition Type:** This field describes the type of the logical partition.

Possible values are:

- Dedicated LPAR
- Shared Processor LPAR

- ▶ **SMT Mode:** This field indicates if Simultaneous Multi-Threading (SMT) is active.  
Possible values are:
  - On
  - Off
- ▶ **Capped:** This field indicates whether a Shared Processor LPAR can exceed its entitlement.  
Possible values are:
  - On
  - Off
- ▶ **Threads:** The number of SMT threads active:
  - For Dedicated LPARs, the number of logical processors equals the number of dedicated physical processors that are assigned to the LPAR, multiplied by the number of SMT threads.
  - For Shared Processor LPARs, the number of logical processors equals the number of virtual processors that are defined for the LPAR, multiplied by the number of SMT threads.
- ▶ **Virtual processors:** This field shows the number of administrator-defined virtual processors. Virtual processors are an abstraction of processors, which are mapped to physical processors in a scheduled manner by the hypervisor. The number of virtual processors defined places an implicit limit on the number of physical processors that can be used by the LPAR.
- ▶ **Capacity Consumed:** This field indicates the actual computing power that the LPAR consumes in units of physical processors.
- ▶ **Guaranteed Capacity:** This field shows the guaranteed physical processor capacity of a Shared Processor LPAR in units of fractional physical processors.
- ▶ **Guaranteed Capacity Consumed:** This field indicates the ratio of the actual consumed physical processor to the entitlement as a percentage. In the case of an uncapped Shared Processor LPAR, the value can exceed 100%.
- ▶ **Available Capacity:** This field indicates the possible available computing power for the LPAR based on the entitlement, which is guaranteed to the LPAR and the current idle capacity in the pool. This value is only available if pool utilization authority is granted.
- ▶ **Available Capacity Busy:** This field indicates the ratio of physical processor user and system ticks that the LPAR consumes to the available capacity for the LPAR as a percentage. This value is only available if pool utilization authority is granted.
- ▶ **Available Capacity Consumed:** This field indicates the ratio of physical processors that the LPAR consumes to the available capacity for the LPAR as a percentage. This value is only available if pool utilization authority is granted.
- ▶ **Additional Capacity Available:** The amount of physical processor capacity that can still be attained by the LPAR.
- ▶ **Weight:** This value is used to determine the allocation of spare pool resources to uncapped SPLPARs.
- ▶ **Capacity Maximum:** The maximum amount of physical processor capacity that can be acquired by the LPAR.

Subsection processor Virtualization Virtual Container presents processor information relevant to and/or defined on the WPAR.

- ▶ **Container Type:** The type of WPAR.

Possible values are:

- Application WPAR
- System WPAR

- ▶ **Container Name:** The WPAR name.

- ▶ **Physical Processors Consumed:** This field indicates the actual computing power that the WPAR consumes in units of physical processors.

Subsection Memory Virtualization Virtual System presents memory information relevant to and/or defined on the LPAR.

- ▶ **Memory Mode:** Indicates whether AME or AMS are active.

Possible values are:

- Dedicated – Neither AME nor AMS is active
- Dedicated Expanded – AME is active, AMS is inactive
- Shared – AME is inactive, AMS is active
- Shared Expanded – Both AME and AMS are active

- ▶ **AME Target Factor:** The ratio between desired expanded memory and true memory.
- ▶ **AME Actual Factor:** The ratio between actual expanded memory and true memory.
- ▶ **AME Expanded Memory:** The effective memory realized by AME.
- ▶ **AME True Memory:** The actual physical memory available to the LPAR.
- ▶ **AME Deficit Memory:** If the desired expanded memory is greater than the actual expanded memory, then the difference is shown.

**Tip:** Refer to the following SAP Note for further information regarding PowerVM and operating system metrics.  
 1131691: Processor utilization metrics of IBM System p®  
 1379340: Enhanced processor utilization metrics on System p

Depending on the partition type and the setting of pool utilization authority (PUA), some of the new processor metrics either are not relevant or cannot be calculated. Table 9-3 summarizes which data is displayed.

Table 9-3 st06 displayed monitoring data

	Dedicated LPAR	SPLPAR with PUA not granted	SPLPAR with PUA granted
Model	✓	✓	✓
Processor	✓	✓	✓
Maximum Processor Frequency	✓	✓	✓
Pool ID		✓	✓
Pool Utilization Authority		✓	✓
Pool processors		✓	✓
Pool processors Idle		✓	✓

	Dedicated LPAR	SPLPAR with PUA not granted	SPLPAR with PUA granted
Pool Maximum Size		✓	✓
Pool Entitled Size		✓	✓
Current Processor Frequency	✓	✓	✓
Partition Name	✓	✓	✓
Partition ID	✓	✓	✓
Partition Type	✓	✓	✓
SMT Mode	✓	✓	✓
Capped		✓	✓
Threads	✓	✓	✓
Virtual processors		✓	✓
Capacity Consumed		✓	✓
Guaranteed Capacity Consumed		✓	✓
Available Capacity			✓
Available Capacity Busy			✓
Available Capacity Consumed			✓
Additional Capacity Available Physical processors idle			✓
Weight		✓	✓
Capacity maximum		✓	✓

The majority of these metrics are concerned with processor utilization monitoring. Although the traditional global processor utilization metrics found in st06 are useful for Dedicated LPARs, they can provide misleading information in the context of Shared Processor LPARs. Although the number of logical processors remains constant, the underlying number of physical processors to which these utilization percentages refer is not constant in the case of a Shared Processor LPAR. For example, in Figure 9-21 on page 135, components for user, system, and I/O wait and idle are identical at 10 am and 12 am. However, the actual physical processor consumption by the LPAR is very different. We cannot simply compare the utilization percentages, as is traditionally done.



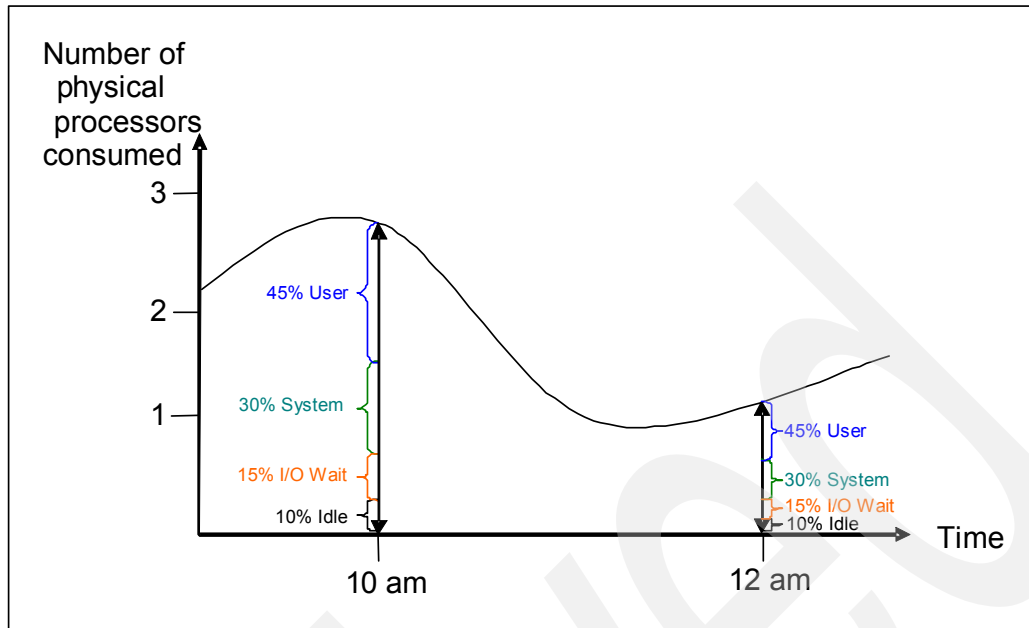


Figure 9-21 Physical processor Consumption versus Global Logical processor Utilization

Likewise, traditional rules of thumb might say that a system is nearing a processor bottleneck if  $\text{user\%} + \text{system\%} > 90\%$  cannot be applied to an SPLPAR.

For an SPLPAR, we need to consider whether more physical processor resources are available for use, as shown in Example 9-19. Additionally, we need to consider whether the SPLPAR can acquire these resources based on configuration and restrictions.

*Example 9-19 Available resources*

---

`available resources = MAX(entitlement, physical processors consumed + pool idle processors)`

`acquirable resources = MIN( entitlement, online virtual processors ) if LPAR is capped or weight=0`  
`= MIN( shared pool processors, online virtual processors ) otherwise`

---

The metrics `available capacity` and `available capacity consumed` encapsulate these considerations and display that the total physical processor resources are available to and can be acquired by a Shared Processor LPAR.

*Example 9-20 Available capacity and available capacity consumed*

---

`available capacity = MIN(available resources, acquirable resources)`

`available capacity consumed = (available capacity / physical processors consumed) * 100`

`available capacity busy = (physical processors(user + system) consumed / available capacity) * 100`

---

Available capacity consumed is based on physical processors consumed, `physc`, while available capacity busy is based on only the user and systems ticks physical processors consumed, `physb`. The difference between `physc` and `physb` is discussed in 8.6.6, "Side effects of PURR based metrics in Shared Pool environments" on page 94.

If the available capacity consumed is less than 100%, the LPAR is not bottlenecked in attaining additional physical processor resources. If available capacity consumed is 100% or near, but available capacity busy is significantly less, the LPAR is still not bottlenecked.

### Special considerations on IBM i

If you are using the new SAP operating system monitor on IBM i, you must consider the following deviations from other operating systems:

- ▶ In the “processor” section, the processor clock rate (referred to as “Current Processor Frequency”) is not displayed for IBM i. You can look up the processor clock rate in *Performance Capabilities Reference*, SC41-0607, which can be found at:  
<http://www.ibm.com/systems/i/advantages/perfmgmt/resource.html>
- ▶ The document lists the processor clock rates based on the processor type, model and feature code. You can find this information in the OS07 output in category “processor Virtualization Host”. Section “Model” is showing the processor type and model, section “Processor” is showing the feature code and serial number. In the processor section under category processor Single, you will not find a breakdown of the processor utilization over the configured processors. The entry for processor 0 will show averages for the processor utilization in all configured processors, while the other entries are showing -1 to indicate that the value is not available.
- ▶ At the time of writing, the *Memory* section in the new operating system monitor did not show data for the configured main storage pools on IBM i. That data is still available in the old transactions ST06, OS06, or OS07. If you are using an SAP release and support package level that routes the transaction ST06 per default to the new operating system monitor, you can still look at the old data by executing the ABAP program RSHOST10 through transaction SE38.

## 9.4.3 SAP Monitoring Infrastructure

The SAP Monitoring Infrastructure (CCMS) consists of several pieces and parts, which include the previously described transaction for the operating system monitor. In this section, we describe all of the pieces of this infrastructure that are relevant for virtualized environments: how to install and properly configure certain parts and how to use integrated features to create reporting (also across systems).

For virtualized environments, the following parts of the SAP Monitoring Infrastructure are of special interest and are described later in detail:

- ▶ A data collector for operating system metrics. This collector is the program saposcol, which is part of every SAP installation.
- ▶ An agent, which can be used to transport the data that saposcol gathers to another system in case the data collector and the monitor for the data are not running on the same system. This agent is the program sapccmsr and is called the CCMS agent.
- ▶ A data repository to store the gathered metrics in the SAP system. The CPH (Central Performance History) is the repository that is used.
- ▶ A transaction to view the gathered metrics. This is the Transaction st06, which we described in 9.4.2, “SAP CCMS operating system monitor” on page 128.
- ▶ An infrastructure to create reporting on the data beyond the available views in transaction st06. Together with the Cst06entral Performance History, the data collectors and the agents that are running on various systems, this infrastructure allows you to create views about the system level and not only for a single LPAR.

The easiest scenario of monitoring virtualized environments with the SAP Monitoring Infrastructure is a single LPAR, which certainly can be also a shared pool LPAR, as shown in Figure 9-22. There are no special configurations necessary to monitor this LPAR, other than a few adjustments that you must make in the CPH if historical data is stored. In this case, you need no CCMS agent because the data that is transferred between saposcol and the monitoring system can occur through a shared memory segment.

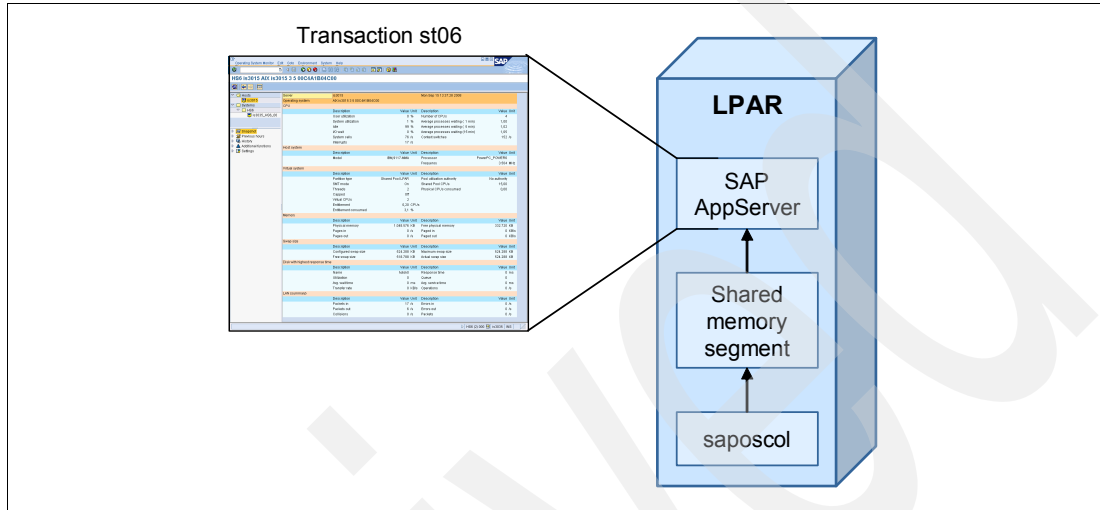


Figure 9-22 SAP monitoring a single LPAR

You can also monitor a set of LPARs with a single monitoring system using the SAP Monitoring Infrastructure. In this case, you have a data collector in every LPAR and a CCMS agent to transfer the gathered metrics to the central monitoring system. Figure 9-23 on page 138 is one example of such an infrastructure.

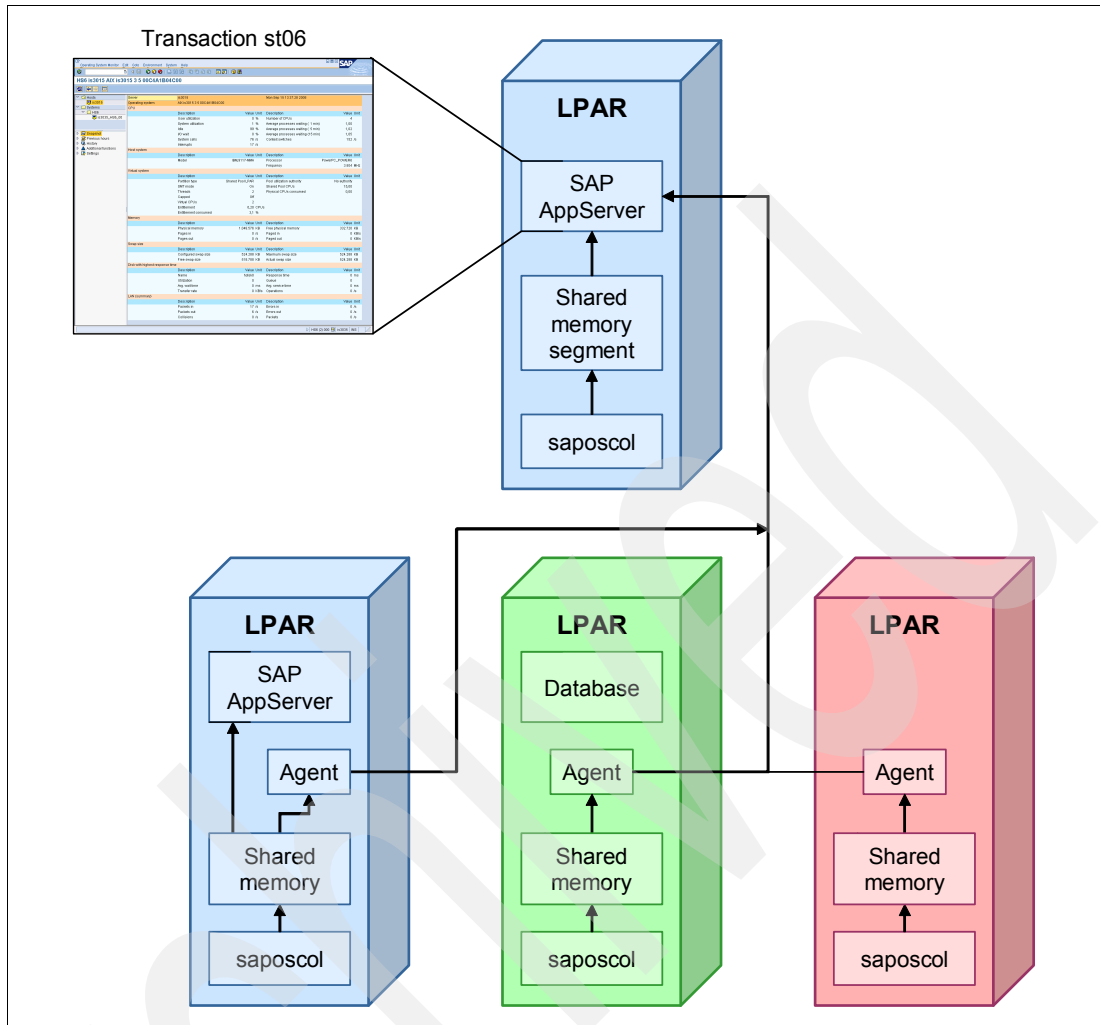


Figure 9-23 SAP monitoring overview

The transaction st06 shows, on the left side of the window, a list of all systems where monitoring data is available. Although you can access this data from multiple systems through one transaction window, you do not get an aggregated view of the available data. Nevertheless the SAP Monitoring Infrastructure provides mechanisms to achieve this goal by using reporting on the CPH. In “Creating reporting on the Central Performance History” on page 153, we describe a simple example of how to utilize this framework. With this functionality, you can create reports that aggregate data across multiple systems.

### saposcol and the CCMS agent

The operating system data collector and the CCMS agent are independent executables. They are not bound to an SAP instance and can also be used to monitor an AIX, IBM i, or Linux partition without an SAP installation. A special case is the Virtual I/O Server because it usually does not allow you to install external software.

In SAP releases prior to 7.10, saposcol and the CCMS agent are usually installed automatically as part of an SAP installation. Beginning with Release 7.10, saposcol is shipped with a separate package called SAPHOSTAGENT. This package is installed together with any SAP installation in Release 7.10 or higher, but it can also be installed separately on a server or partition with older SAP releases or without any SAP instance.

Because of the restrictions in the Virtual I/O Server, we discuss the installation of saposcol and the CCMS agents on non-VIOS partitions and VIOS partitions separately.

### ***Installing in a non-VIOS LPAR***

#### **Step 1:** Installing saposcol and the CCMS agent

1. Download and install the SAPCAR tool from the SAP Software Distribution Center. For AIX and Linux, follow the instructions in SAP Note 212876. For IBM i, follow the instructions in SAP Note 863821.
2. Follow the instructions in the attachment of SAP Note 1031096 to download and install the SAPHOSTAGENT package. This creates the directory `/usr/sap/hostctrl` with subdirectories `exe` (for the executables) and `work` (for traces, log files and other data). It will also create the user `sapadm` (IBM i: `*USRPRF SAPADM`) that can be used to administer saposcol and the CCMS agent.
3. Download the CCMAGENT package from the SAP Software Distribution Center and rename it to `ccmagent.sar`. On AIX and Linux you can decompress the archive with the command:  

```
sapcar -xvf ccmagent.sar
```

  
On IBM i, follow the instructions in SAP Note 1306787. Copy the `sapccmsr` executable to directory `/usr/sap/hostctrl/exe`.

#### **Step 2:** Starting saposcol

If you follow the installation instructions in the attachment of SAP Note 1031096, SAPHOSTAGENT and saposcol along with it will be started automatically at each IPL or reboot of the partition. If you need to start saposcol manually, follow these steps:

On AIX and Linux:

1. Switch to the `sapadm` user that you created during the previous installation steps:  

```
# su - sapadm
```
2. Change to the path where saposcol resides:  

```
# cd /usr/sap/hostctrl/exe
```
3. Launch saposcol:  

```
# saposcol -l
```

On IBM i:

1. Sign on with a user profile that has `*JOBCTL` and `*ALLOBJ` authorities.
2. Launch saposcol by calling the executable:  

```
CALL PGM(R3SAP400/SAPOSCOL) PARM('-1')
```

#### **Step 3:** Configuring and starting the CCMS agent

An SAP user with sufficient RFC permissions is required to run the CCMS agent. The easiest way to create such a user is to use the transaction RZ21 and create the CSMREG user on the SAP instance that you use for monitoring, as shown in Figure 9-24 on page 140.

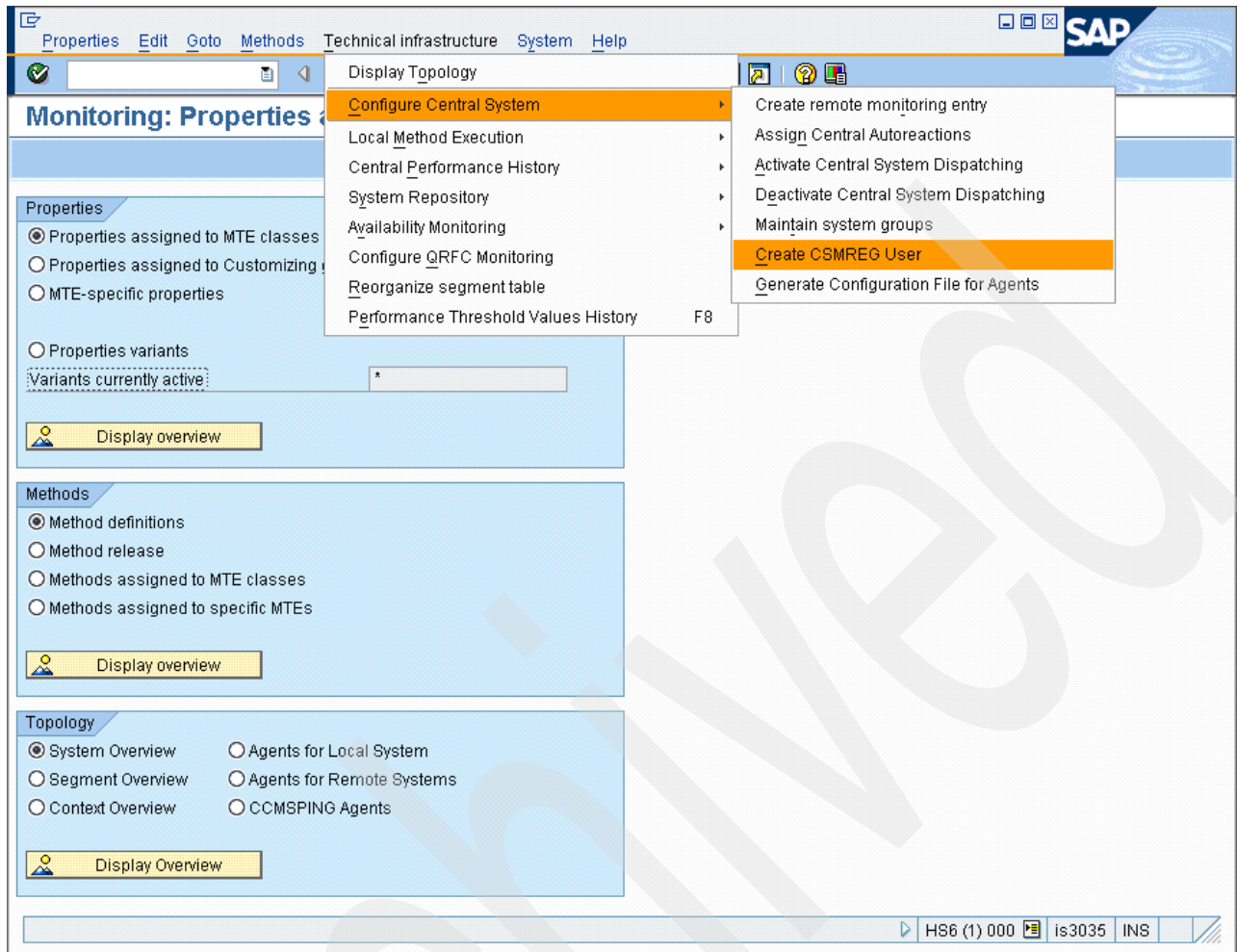


Figure 9-24 Create CSMREG user from RZ21<sup>7</sup>

In the next panel enter the desired password, and press Enter. The CSMREG user is created, as shown in Figure 9-25 on page 141.

<sup>7</sup> Copyright SAP AG 2008. All rights reserved.

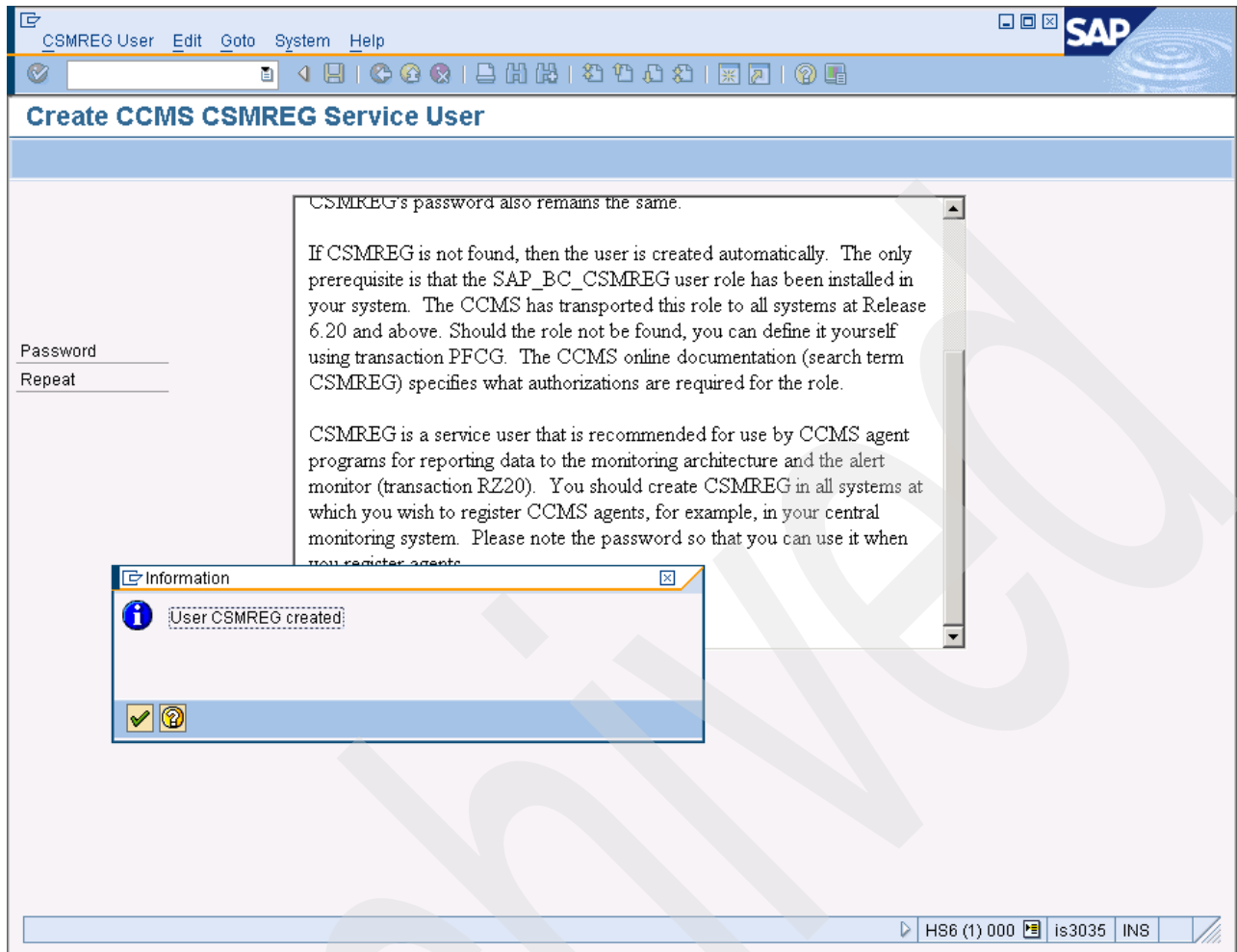


Figure 9-25 CSMREG user created<sup>8</sup>

Besides the CSMREG user, another SAP user with admin privileges on the central monitoring system is required to complete the CCMS configuration step. This user can be the user DDIC or any other user with administration privileges.

To configure the CCMS agent the following information is needed:

- ▶ Host name, SAP system ID, and client number for each central system to which the agent reports
- ▶ An SAP user with Admin privileges and password
- ▶ The CSMREG user and password
- ▶ Language used
- ▶ Host name of application server
- ▶ System number (00 - 98)
- ▶ Route string [optional]
- ▶ Trace level [0]
- ▶ Gateway information if not the default

<sup>8</sup> Copyright SAP AG 2008. All rights reserved.

To start the CCMS agent registration and configuration on AIX or Linux, enter the following commands:

```
# su - sapadm
# cd /usr/sap/hostctrl/exe
# sapccmsr -R
```

To start the CCMS agent registration and configuration on IBM i, sign on with a user profile that has \*JOBCTL and \*ALLOBJ special authorities and enter the following commands:

```
CD DIR('/usr/sap/hostctrl/exe')
CALL PGM(R3SAP400/SAPCCMSR) PARM('-R')
```

The configuration is straightforward when you use predefined options and information. In the rare case of a wrong configuration, the easiest thing to do is to remove the configuration file /usr/sap/tmp/sapccmsr/csmconf and run **sapccmsr** in configuration mode again.

To start the agent automatically after each restart of an AIX or Linux partition, insert the following command into /etc/inittab:

```
/usr/sap/hostctrl/exe/sapccmsr -DCCMS
```

To start the agent automatically after each restart of an IBM i partition, enter the following commands to the system startup program that is defined in system value QSTRUPPGM:

```
CD DIR('/usr/sap/hostctrl/exe')
CALL PGM(R3SAP400/SAPCCMSR) PARM('-DCCMS')
```

### **Installing in a VIOS LPAR**

Being an appliance partition, the VIOS requires careful installation treatment. For this reason, a specially created “saposcol for vios” package has been put together to assist installation. This package has been explicitly tested and approved by IBM to be installed on the VIOS.

**Tip:** Refer to the following further details and support information

- ▶ SAP OSS note 1379855: Installation of saposcol and sapccmsr on IBM VIOS
- ▶ IBM white paper: “Include VIOS Partitions into SAP Performance Monitoring”:  
<http://www.ibm.com/support/techdocs/atsmastr.nsf/WebIndex/WP101541>

### **Step 1: Installing saposcol and the CCMS agent**

1. Download the appropriate install package, sapviosagent.sar, from SAP Service Marketplace. For example, the 7.20 package can be on SAP Service Marketplace at:

```
Support Packages and Patches → ? A-Z Index → N → SAP NETWEAVER → SAP EHP2 FOR
SAP NETWEAVER 7.0 → Entry by Component → Application Server ABAP
→ SAP KERNEL 7.20 64-BIT → AIX 64Bit → #Database independent
```

2. Unpack the SAP install package to get the AIX bff-install.p file

```
# SAPCAR -xvf sapviosagent.sar
SAPCAR: processing archive sapviosagent.sar (version 2.00)
x ./tmp/sapviosagent.com.sap.7.20.0.97.bff
SAPCAR: 1 file(s) extracted
```

3. Transfer the bff file to the VIO Server using ftp.

```
# ftp is314v1
Connected to is314v1.wdf.sap.corp.
220 is314v1 FTP server (Version 4.2 Tue Sep 14 20:17:37 CDT 2010) ready.
```



```

Name (is314v1:root): padmin
331 Password required for padmin.
Password:
230-Last unsuccessful login: Tue Jul 19 13:14:18 DFT 2011 on ssh from
is3089.wdf.sap.corp
230-Last login: Tue Jul 19 16:41:59 DFT 2011 on ftp from is3110.wdf.sap.corp
230 User padmin logged in.
ftp> bin
200 Type set to I.
ftp> put sapviosagent.com.sap.7.20.0.97.bff
200 PORT command successful.
150 Opening data connection for sapviosagent.com.sap.7.20.0.97.bff.
226 Transfer complete.
12082688 bytes sent in 0.298 seconds (3.96e+04 Kbytes/s)
local: sapviosagent.com.sap.7.20.0.97.bff remote:
sapviosagent.com.sap.7.20.0.97.bff
ftp> quit
221 Goodbye.

```

4. Log in to the VIOS as user padmin and obtain a root shell by executing the following command:

```

$ ssh padmin@is314v1
padmin@is314v1's password:
Last unsuccessful login: Tue Jul 19 13:14:18 DFT 2011 on ssh from
is3089.wdf.sap.corp
Last login: Tue Jul 19 16:42:58 DFT 2011 on ftp from is3110.wdf.sap.corp
$ oem_setup_env

```

5. Install the bff file

```

# inutoc .
# installp -acXY -d . sapviosagent.com.sap
+-----+
Pre-installation Verification...
+-----+
Verifying selections...done
Verifying requisites...done
Results...

SUCSESSES
-----
Filesets listed in this section passed pre-installation verification
and will be installed.

Selected Filesets
-----
sapviosagent.com.sap 7.20.0.97                # Installation saphostcontrol

<< End of Success Section >>

+-----+
BUILDDATE Verification ...
+-----+
Verifying build dates...done
FILESET STATISTICS
-----

```

```
1 Selected to be installed, of which:
  1 Passed pre-installation verification
-----
1 Total to be installed
```

```
+-----+
                        Installing Software...
+-----+
```

```
installp: APPLYING software for:
          sapviosagent.com.sap 7.20.0.97
```

```
group sapsys does not exist
user sapadm does not exist
user sapadm not member of group sapsys
```

```
total 0
-rwxr-xr-x   1 1459   15      3059109 Jul 18 22:11
/usr/sap/hostctrl/exe/saposcol
13:42:23 19.07.2011    LOG: Using PerfDir (DIR_PERF) = /usr/sap/tmp
*****
* This is Saposcol Version COLL 21.03 720 - AIX v12.51 5L-64 bit 110419
* Usage:  saposcol -l: Start OS Collector
*          saposcol -k: Stop OS Collector
*          saposcol -d: OS Collector Dialog Mode
*          saposcol -s: OS Collector Status
* Starting collector (create new process)
*****
13:42:30 19.07.2011    LOG: Using PerfDir (DIR_PERF) = /usr/sap/tmp
INFO: New Saposcol collection interval is 60 sec.
Finished processing all filesets. (Total time: 8 secs).
```

```
+-----+
                        Summaries:
+-----+
```

The installation of the package automatically does the following:

- ▶ Adds the necessary user and group.
- ▶ Creates the appropriate directory paths and places the saposcol and sapccmsr binary in those paths.
- ▶ Starts the saposcol process.

### Step 2: Configuring and starting the CCMS agent

This step is identical to Step 3 of “Installing in a non-VIOS LPAR” on page 139.

### Configuring the RFC connection for the CCMS agent

To monitor operating system metrics in a distributed virtualized environment, you must transfer the data that saposcol gathers and the CCMS agent to the central monitoring system. The underlying technology is based on RFC connections. Therefore, configure those RFC connections in advance, as shown in Figure 9-26 on page 145.

To configure the RFC connection for the CCMS agent:

1. Start the transaction AL15 on the central monitoring system.
2. Insert the saposcol-Destination on the remote LPAR: SAPCCMSR.<Hostname>.99.
3. To modify a destination, press **Modify**.

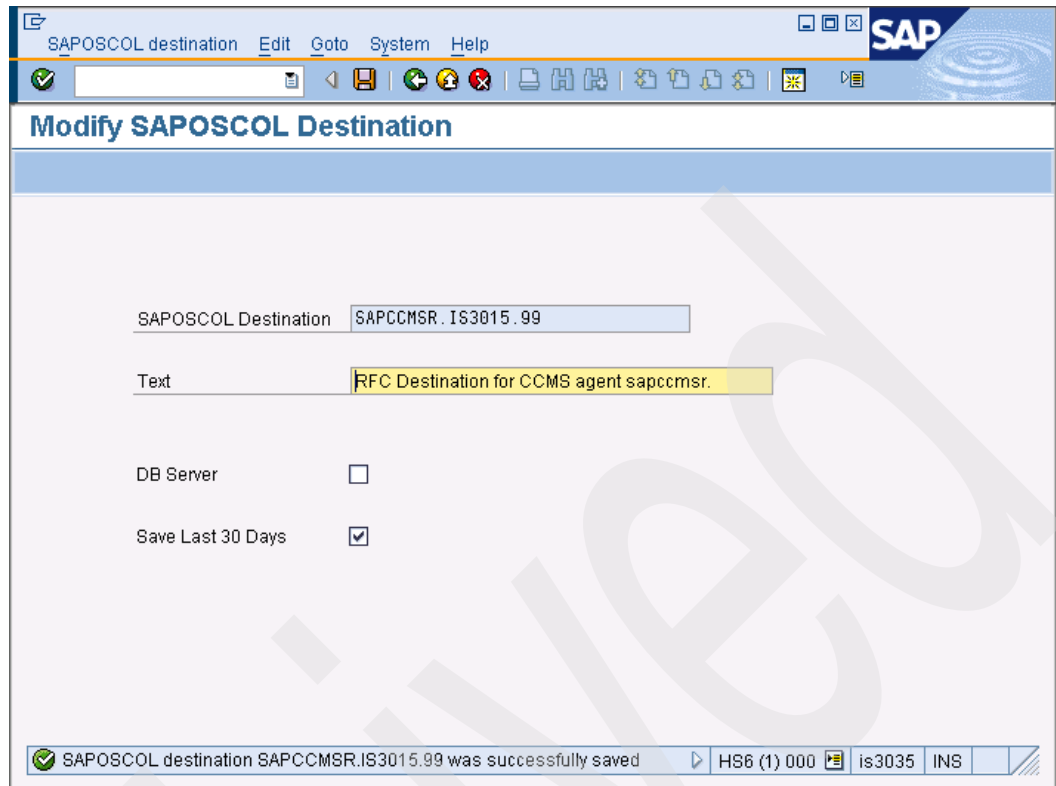


Figure 9-26 *Modify the saposcol destination*<sup>9</sup>

Now the remote LPAR monitoring information is visible, for example in the transaction st06, as shown in Figure 9-27 on page 146. There we added the RFC connection for the system is3015.

<sup>9</sup> Copyright SAP AG 2008. All rights reserved.

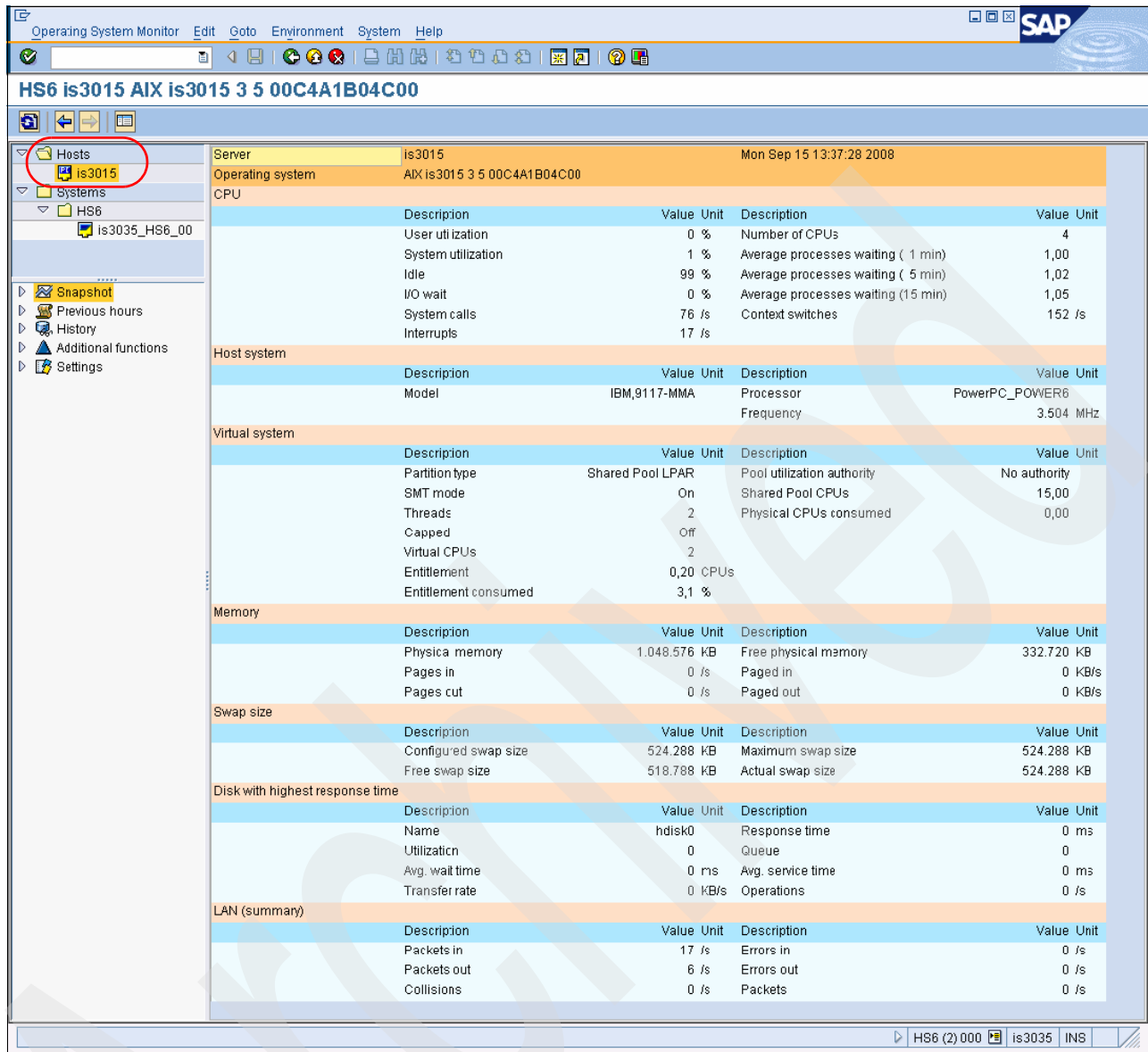


Figure 9-27 Report of a remote LPAR in st06<sup>10</sup>

## Activating and customizing the Central Performance History

The Central Performance History (CPH) stores historical data. Before you can use this functionality, you must activate and configure the CPH. In this example, we use, in most cases, the transaction RZ21, but also transaction RZ23n is a valid entry point into the CPH functions.

To activate the CPH call transaction RZ21:

1. Select **Technical Infrastructure** → **Central Performance History** → **Background Jobs**, as shown in Figure 9-28 on page 147.

<sup>10</sup> Copyright SAP AG 2008. All rights reserved.

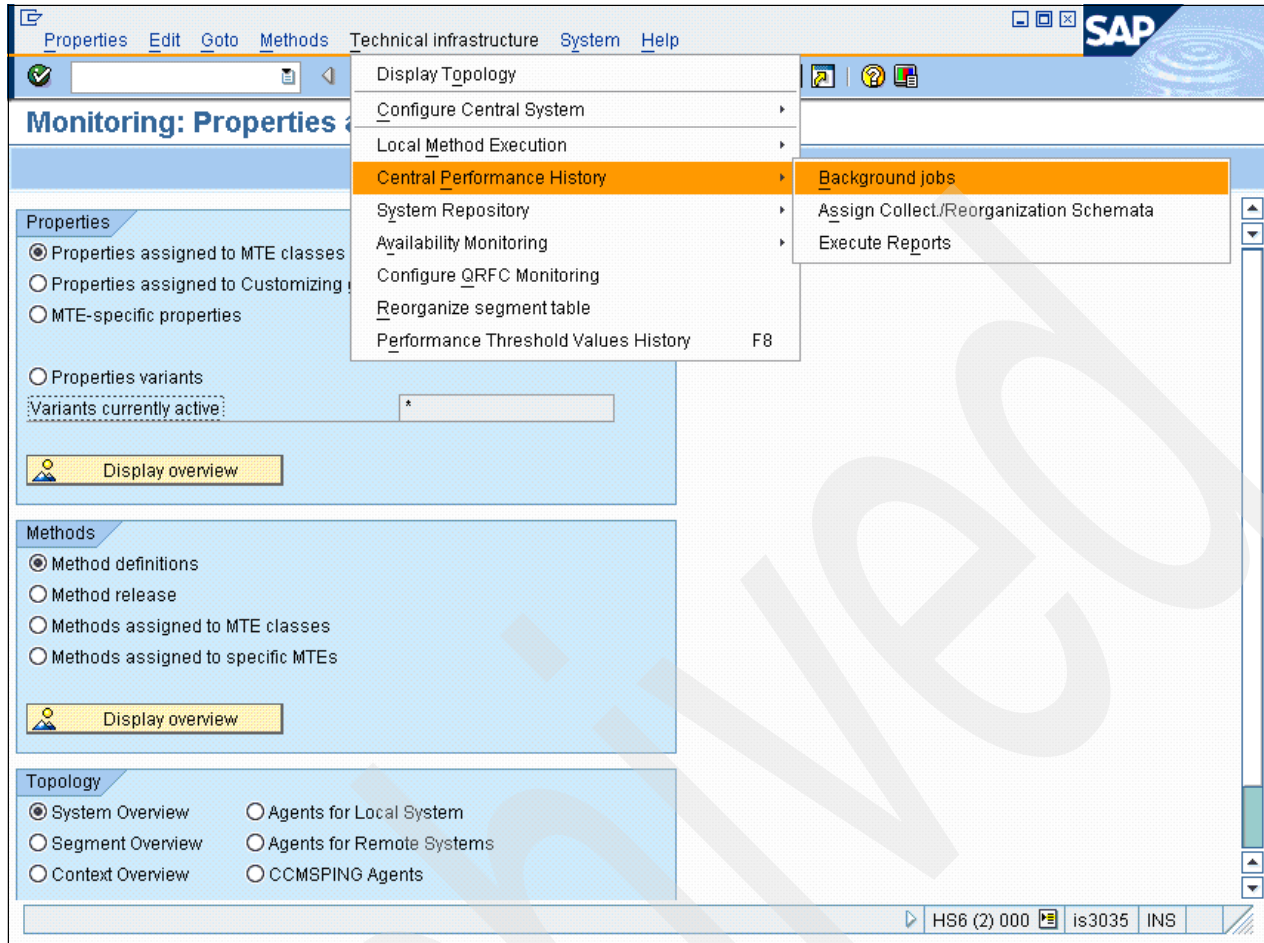


Figure 9-28 Activate CPH from RZ21<sup>11</sup>

2. If jobs are already scheduled, a list with those jobs is displayed; otherwise, confirm the activation of the CPH.

After this basic activation, you can define the granularity of data collection. SAP provides on [help.sap.com](http://help.sap.com), which has detailed instructions about what you can do beyond the example that we give in this book:

[http://help.sap.com/saphelp\\_nw2004s/helpdata/en/c2/49773dd1210968e1000000a114084/content.htm](http://help.sap.com/saphelp_nw2004s/helpdata/en/c2/49773dd1210968e1000000a114084/content.htm)

We are following the proposed steps from SAP to give a simple example:

1. Define how long and with which granularity performance data must be retained in the CPH.
2. Select the performance metrics to be stored in the CPH.

### **Defining the length and granularity of data that is stored in the CPH**

This setting is defined in the Collection and Reorganization Schemata. SAP delivers predefined schemata that meets most requirements. The schemata also contains information about the calculation rules that are to be used to aggregate information. When you do this, you can weight the different hours of the day and also different days differently, to hide public holidays or hours during the night during the calculation of the aggregates.

<sup>11</sup> Copyright SAP AG 2008. All rights reserved.

In this example, as shown in Figure 9-29, we show how to define a day and a collection schema:

1. We start to call the transaction RZ21 and select **Technical infrastructure** → **Central Performance History** → **Assign Collect/Reorganization Schemata**.

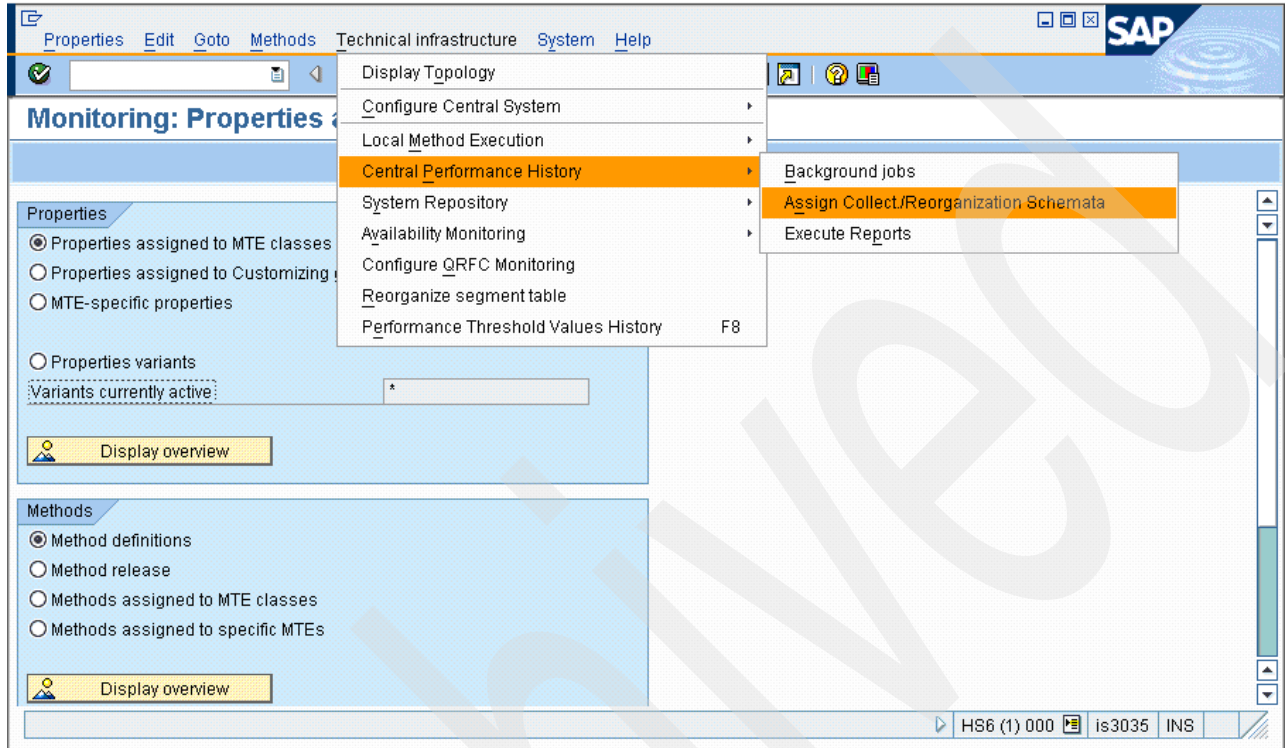


Figure 9-29 SAP GUI menu path<sup>12</sup>

2. In the next transaction window, select **Goto** → **Collection/Reorganization Schema**, and switch to the Advanced Configuration window, as shown in Figure 9-30. An alternative path calls RZ23n, and select **Maintain Schemas**.

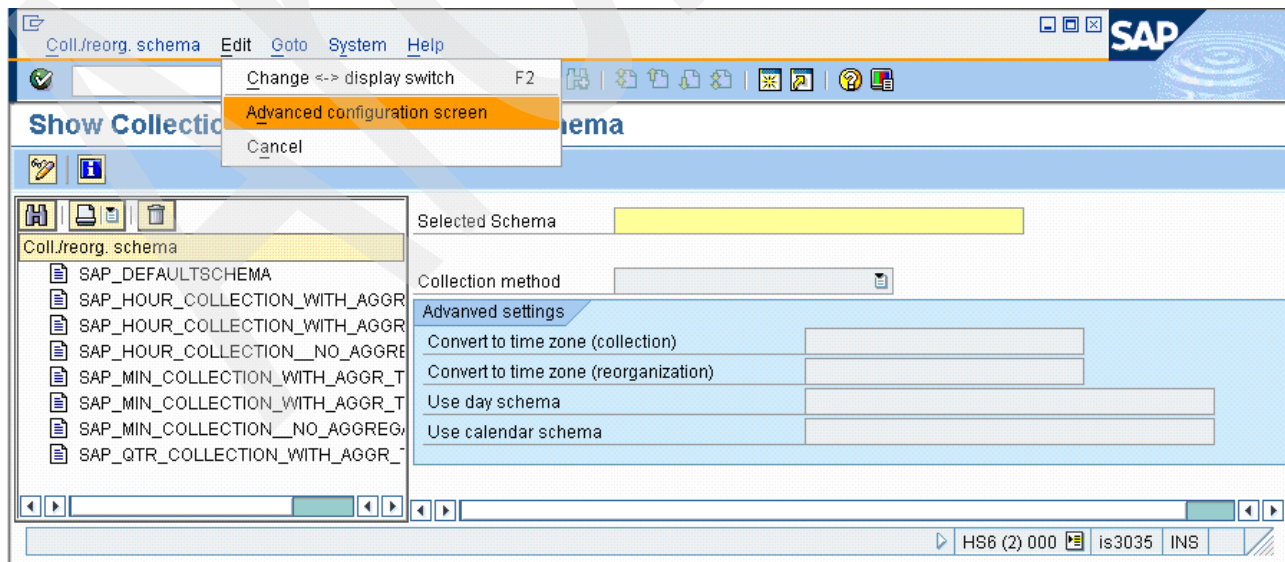


Figure 9-30 SAP GUI menu path<sup>13</sup>

<sup>12</sup> Copyright SAP AG 2008. All rights reserved.

- We define our own SAPOSCOL\_DAY schema to prioritize certain hours in the day, as shown in Figure 9-31 and Figure 9-32 on page 150.

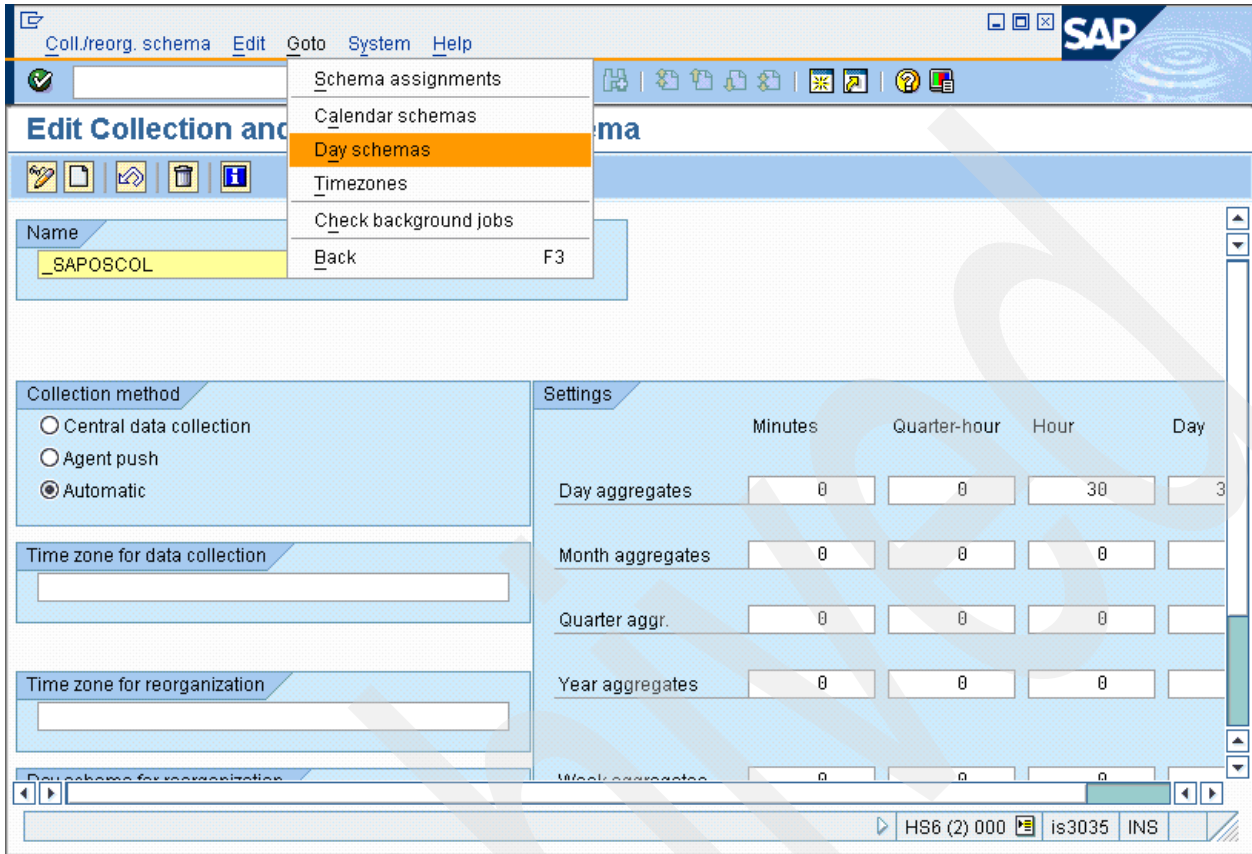


Figure 9-31 Switch to the day schema configuration window<sup>14</sup>

<sup>13</sup> Copyright SAP AG 2008. All rights reserved.

<sup>14</sup> Copyright SAP AG 2008. All rights reserved.



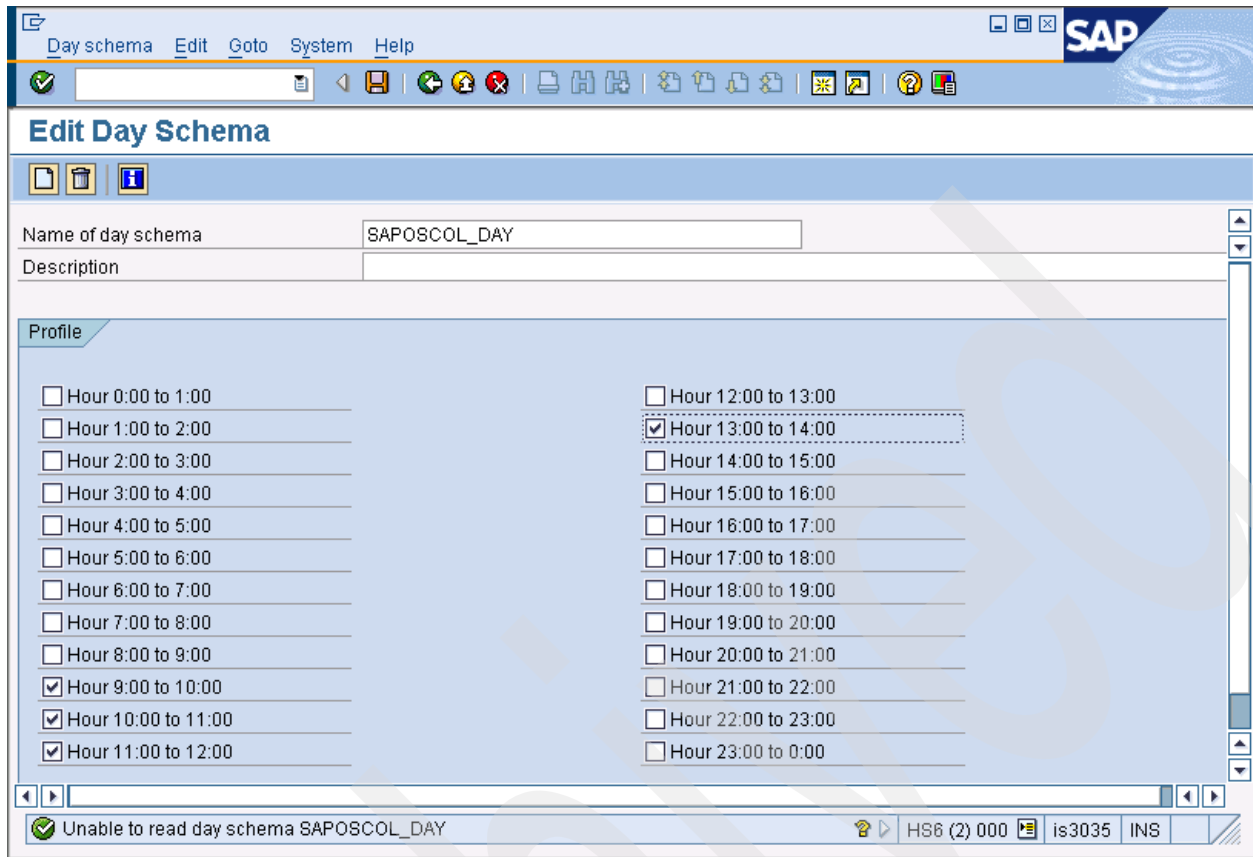


Figure 9-32 Define which hours of the day the schema is active<sup>15</sup>

- In the next step, we can use that day schema to define our collection schema, as shown in Figure 9-33 on page 151.

<sup>15</sup> Copyright SAP AG 2008. All rights reserved.



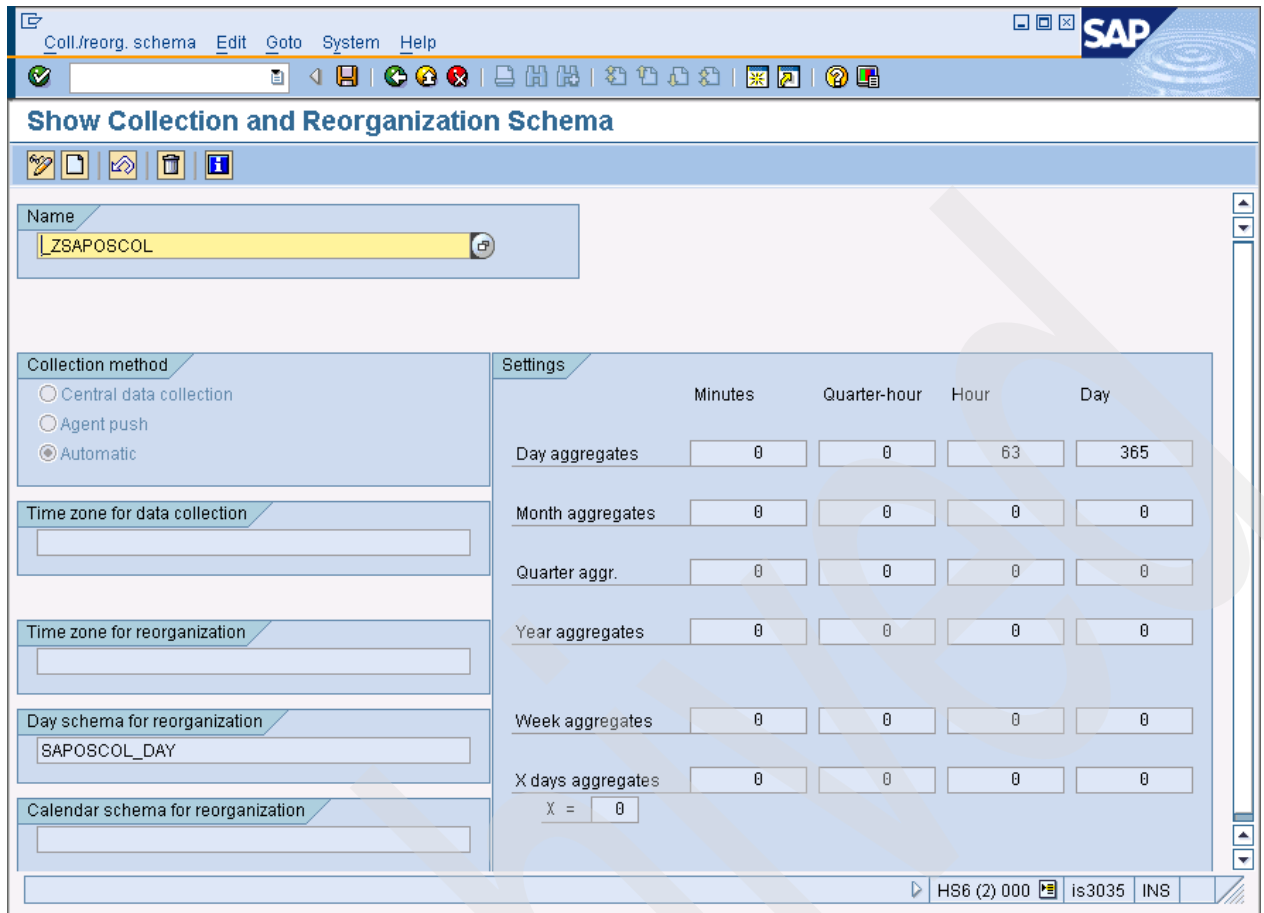


Figure 9-33 Advanced configuration window for `_ZSAPOSCOL` collection schema<sup>16</sup>

### Selecting the performance metrics that are to be stored in the CPH

You can assign collection and reorganization schemata to any MTE classes of any systems. For further information, refer to Assigning Collection and Reorganization Schemata to Performance Values at:

[http://help.sap.com/saphelp\\_nw2004s/helpdata/en/04/4d773dd1210968e1000000a114084/content.htm](http://help.sap.com/saphelp_nw2004s/helpdata/en/04/4d773dd1210968e1000000a114084/content.htm)

MTE classes are the mechanism to access metrics in CCMS. To select the metrics:

1. Go either back to the transaction RZ21, and select **Technical infrastructure** → **Central Performance History** → **Assign Collect/Reorganization**, or call RZ23n, and choose **Assign Procedure**.

In the given example, as shown in Figure 9-34 on page 152, we collect for the non-SAP LPAR running our agent and collector (System ID column) the processor- and AIX-specific virtualization metrics (MTE Classes column).

2. Select the intervals in which the data is aggregated, in the Collection/Reorganization Schema.

<sup>16</sup> Copyright SAP AG 2008. All rights reserved.

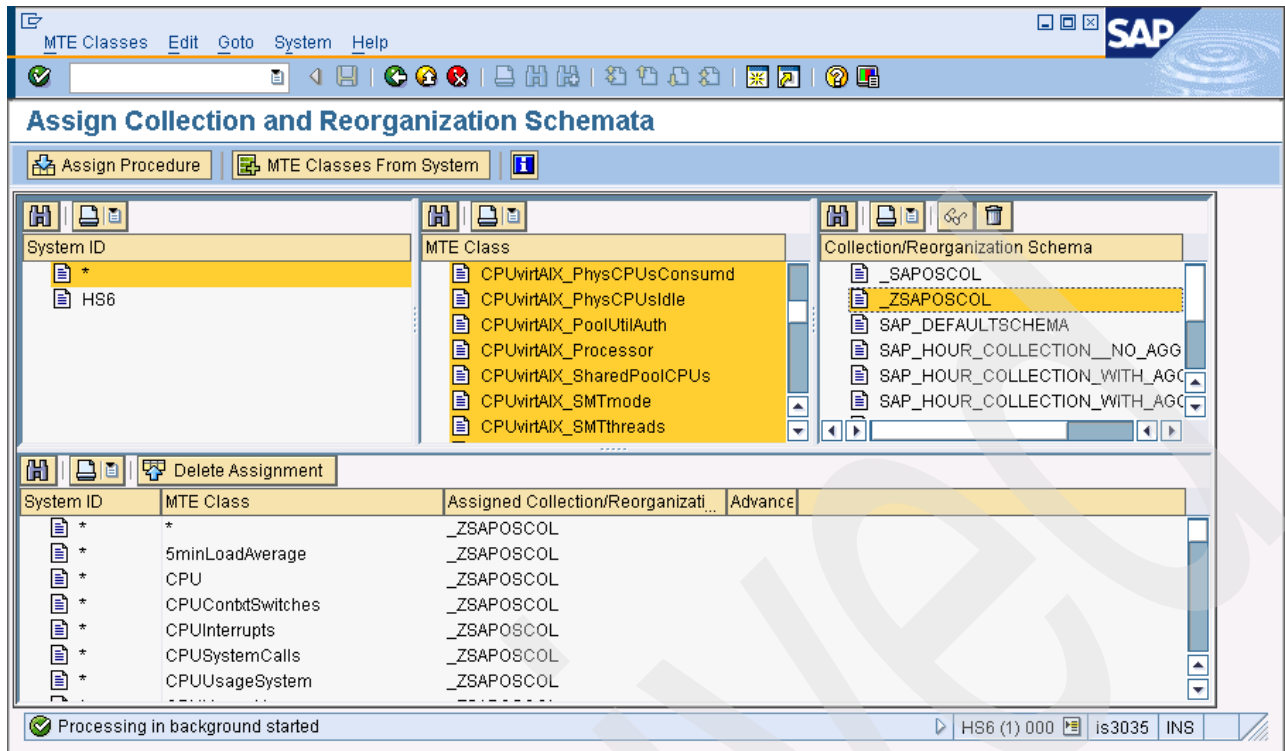


Figure 9-34 Define the metrics to collect in the CPH from the remote LPAR<sup>17</sup>

Now the system starts to gather historical data. Depending on the configuration of the schemas in the first step, the CPH is filled faster or slower. Figure 9-35 on page 153 shows how the historical data is presented in st06.

<sup>17</sup> Copyright SAP AG 2008. All rights reserved.

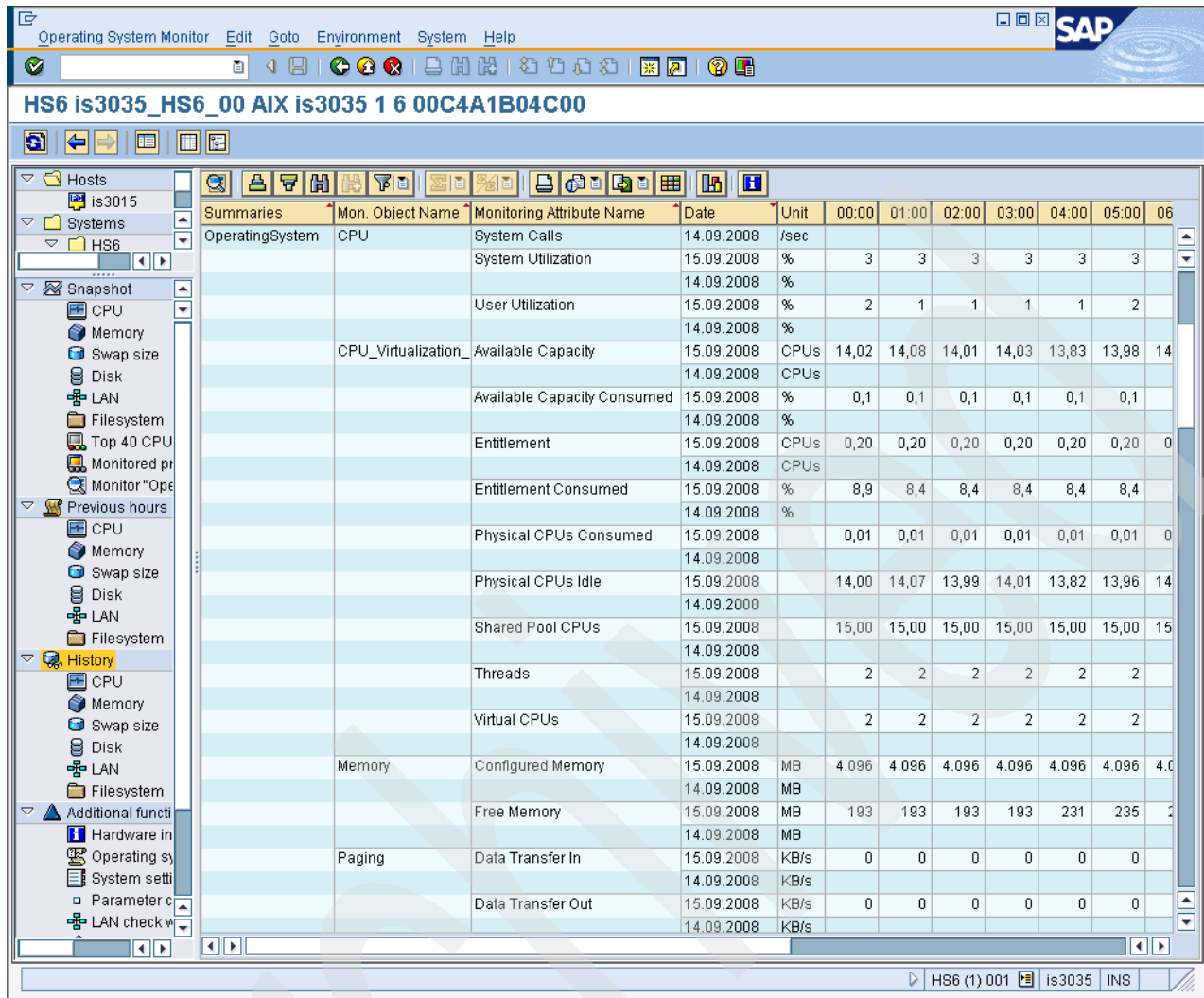


Figure 9-35 CPH data in st06<sup>18</sup>

## Creating reporting on the Central Performance History

After historical data is available in the CPH, you can use the integrated reporting framework from the SAP Monitoring Infrastructure. In the following two procedures, we describe how to use this functionality. On help.sap.com, SAP provides more detailed instructions:

[http://help.sap.com/saphelp\\_nw2004s/helpdata/en/c2/49773dd1210968e1000000a114084/content.htm](http://help.sap.com/saphelp_nw2004s/helpdata/en/c2/49773dd1210968e1000000a114084/content.htm)

### Creating the report definitions

A report definition has two steps:

- ▶ Selecting the MTE classes
- ▶ Selecting the aggregation method, which implicitly defines the layout of the presentation view.

To select the MTE classes:

1. Switch to transaction RZ23n, and choose **Define Reports**.

<sup>18</sup> Copyright SAP AG 2008. All rights reserved.

2. Generate a report giving a cross partition view of the physical processor consummation. In our system, we have one shared pool with 15 processors and only two partitions (one LPAR with and another without an SAP instance).
3. In the Edit Report Definition window, define a new report called CROSS\_PAR\_PHYSprocessor\_CONSUMED (name). We defined two groups. The group Phys, processors consumed summarizes the consumed processors of the two LPARs, whereas the group Shared Pool only shows the static value of how many processors are assigned to that pool, as shown in Figure 9-36. Here, we assume that both LPARs are running on the same pool.

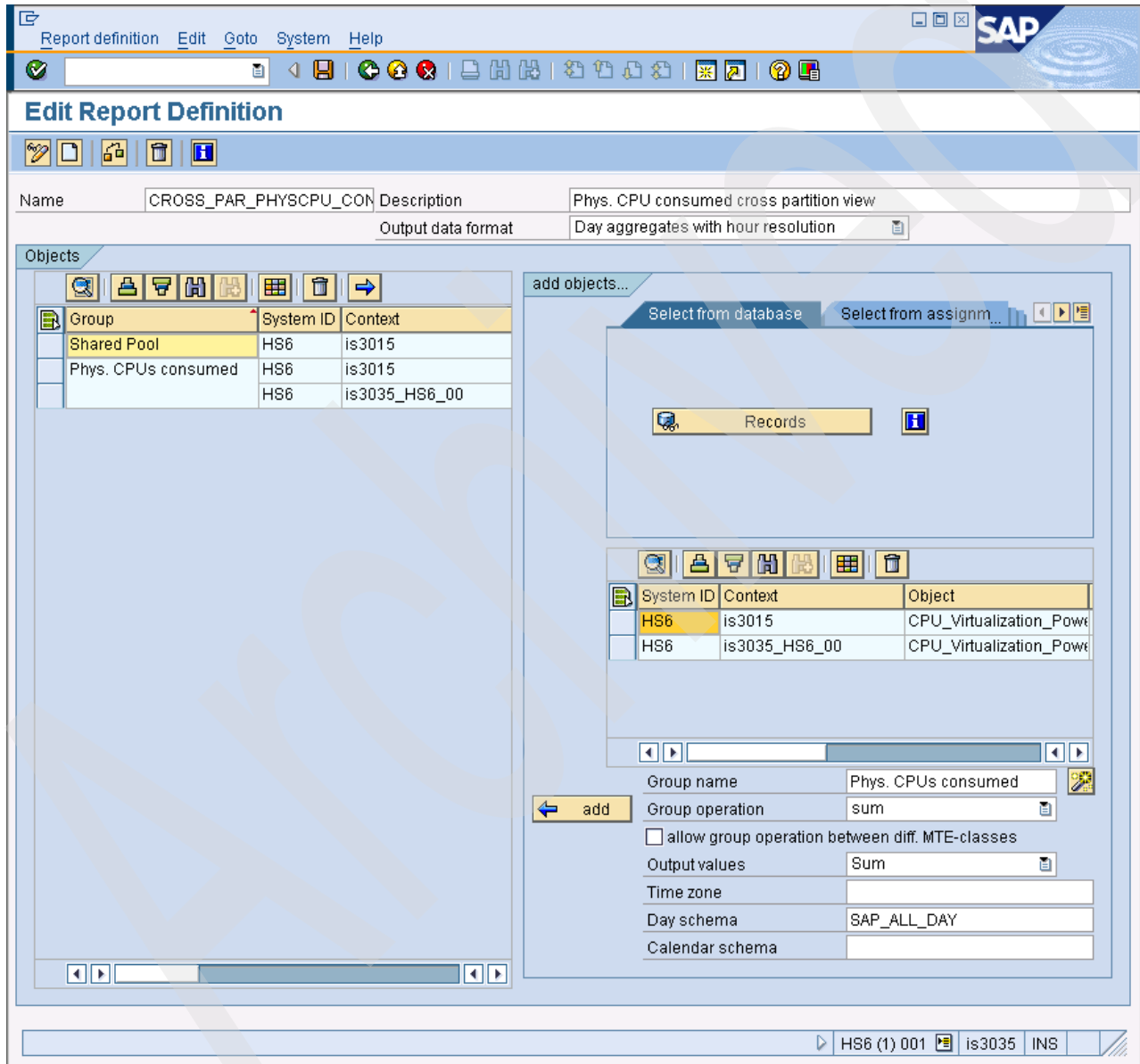


Figure 9-36 Report editor in RZ23n<sup>19</sup>

4. Save the created report.

<sup>19</sup> Copyright SAP AG 2008. All rights reserved.

### Executing the report

You can execute the defined report directly or schedule it as a regular job. Depending on its settings, you can display the report directly on the window or redirect the output to a file:

Return to transaction RZ23n, and choose **Schedule and Execute Report**. As shown in Figure 9-37, we target the output to what is only possible by our chosen method to execute the report directly by pressing **Direct Execution**.

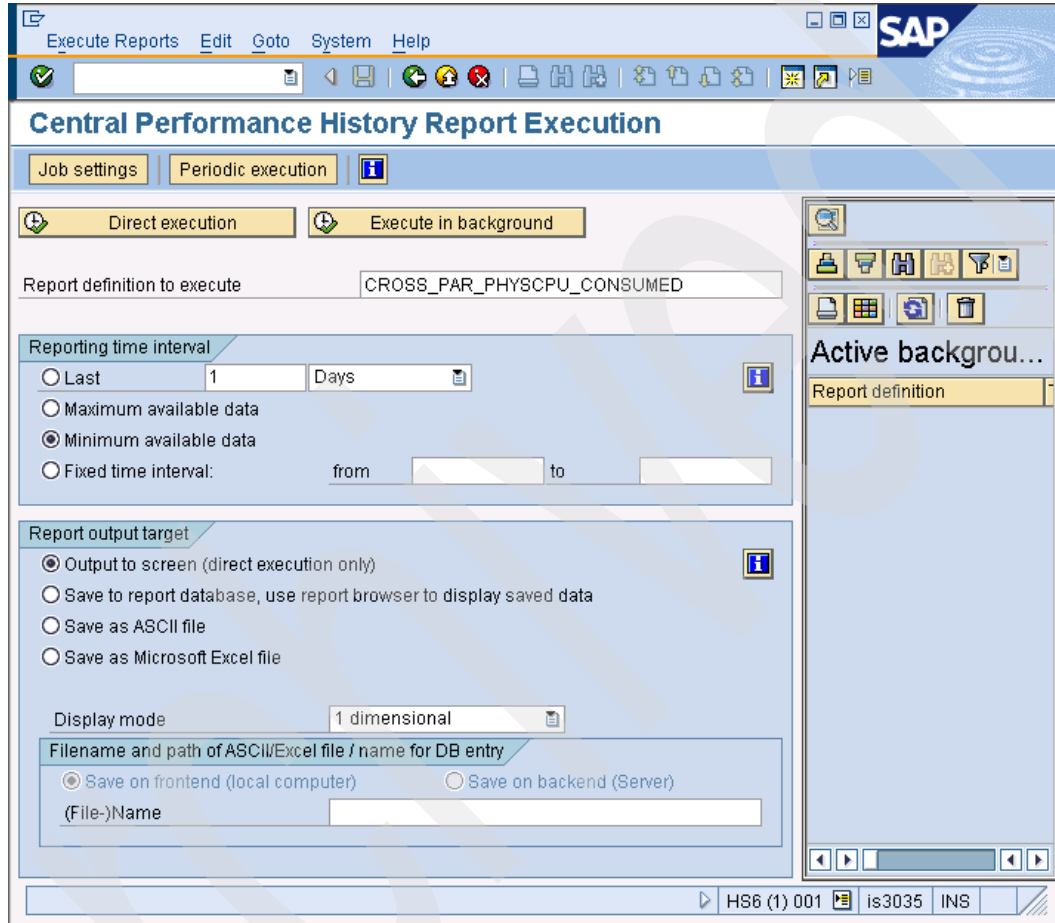


Figure 9-37 Transaction RZ23n; execute and schedule report window<sup>20</sup>

The output window for the report provides additional functionality to draw graphs or further data aggregation, such as summarizing columns, as shown in Figure 9-38 on page 156.

<sup>20</sup> Copyright SAP AG 2008. All rights reserved.

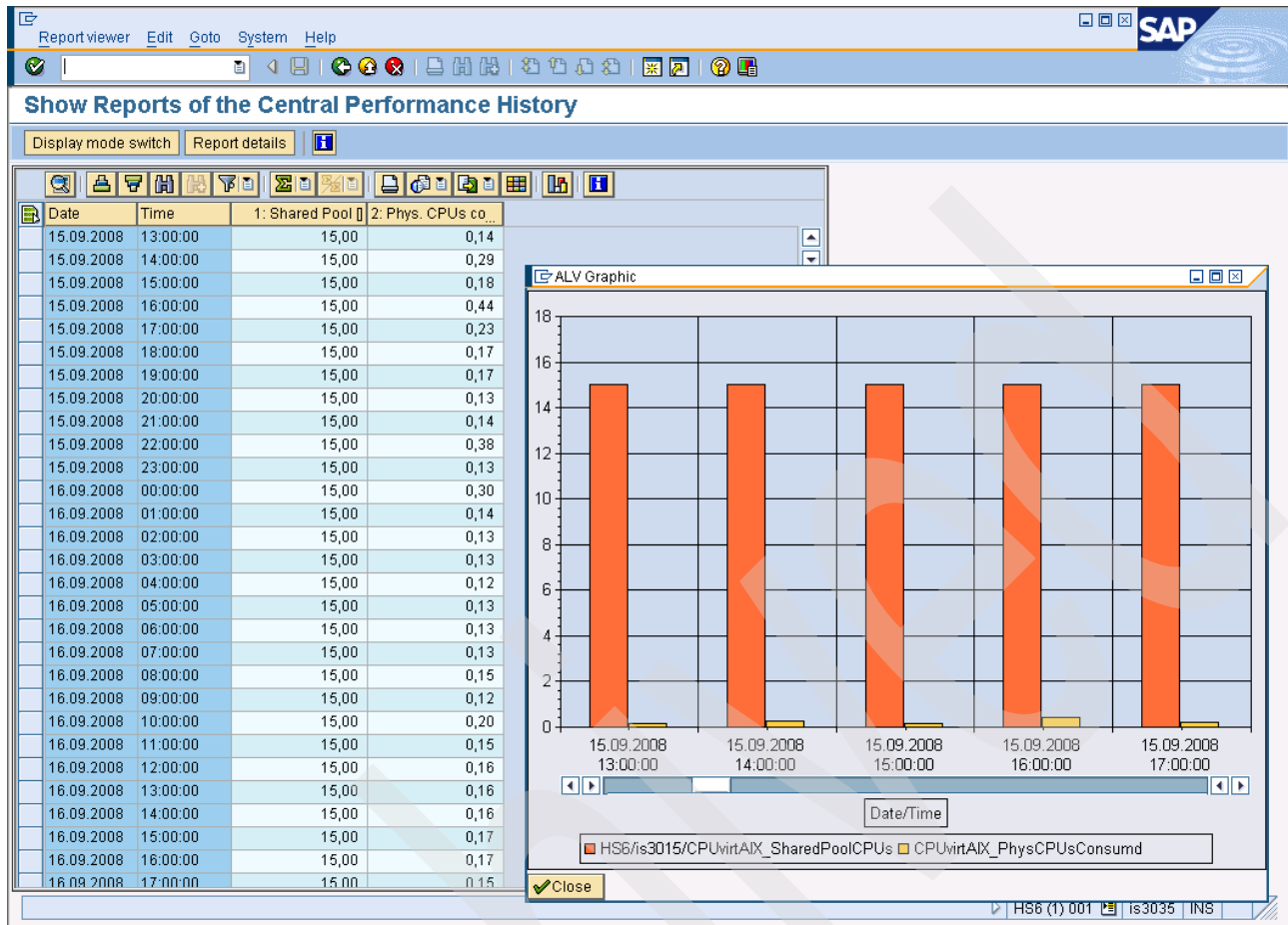


Figure 9-38 Example output of the generated report<sup>21</sup>

## 9.5 SAP EarlyWatch and SAP GoingLive check services

The SAP GoingLive and EarlyWatch services are proactive services during the implementation phase and production phase, respectively, that provide SAP system analysis and recommendations to ensure performance and availability. These services are available through SAP.

More information is available at the SAP Service Marketplace at:

<http://service.sap.com/earlywatch>  
<http://service.sap.com/goinglive>

The analysis provided by these services was updated to take into account the different LPAR configuration possibilities on IBM Power Systems. You can see the virtualization analysis and details in the session details using the Solution Manager.

<sup>21</sup> Copyright SAP AG 2008. All rights reserved.

## 9.6 IBM Insight for SAP

IBM Insight for SAP is an IBM offering that provides customers with an in-depth report that details the SAP workload/performance statistics and host resource utilization. It consists of a downloadable data collector utility, the Insight Collector, and a report generation service by IBM. Both the utility and the subsequent report generation service are provided free of charge by IBM America's Advanced Technical Support (ATS) and Techline organizations. The report generation is limited to production systems and also can only be requested once per quarter per SAP SID.

The latest version, IBM Insight for SAP Version 5, was expanded to include metrics that are relevant to PowerVM virtualization, for example, for a Shared Processor LPAR, the report includes graphs over a measured period of time for the following data:

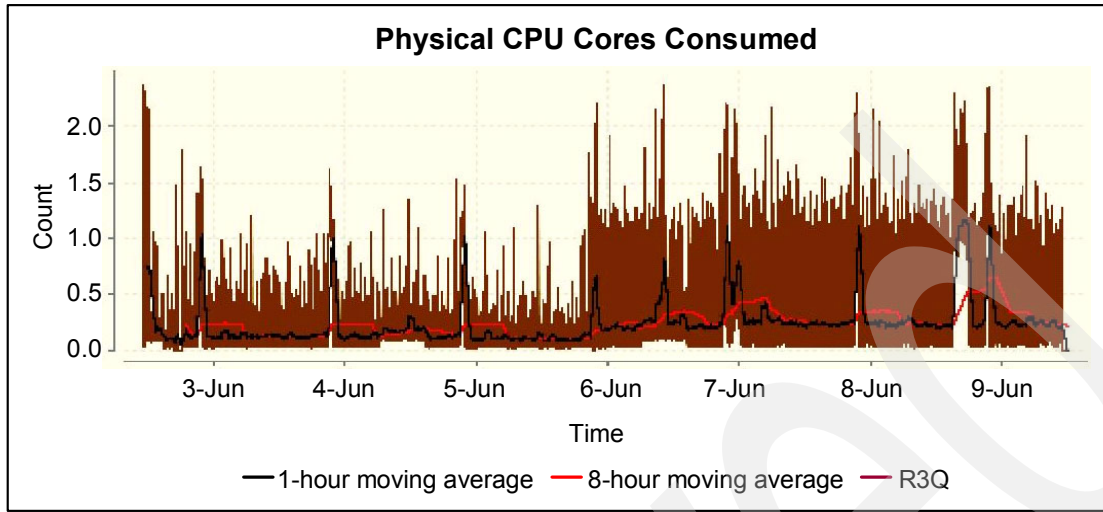
- ▶ Physical processor cores consumed
- ▶ Physical processor cores idle
- ▶ Percent of processor entitlement consumed
- ▶ Physical processor entitlement
- ▶ Physical processor available

Features of Version 5:

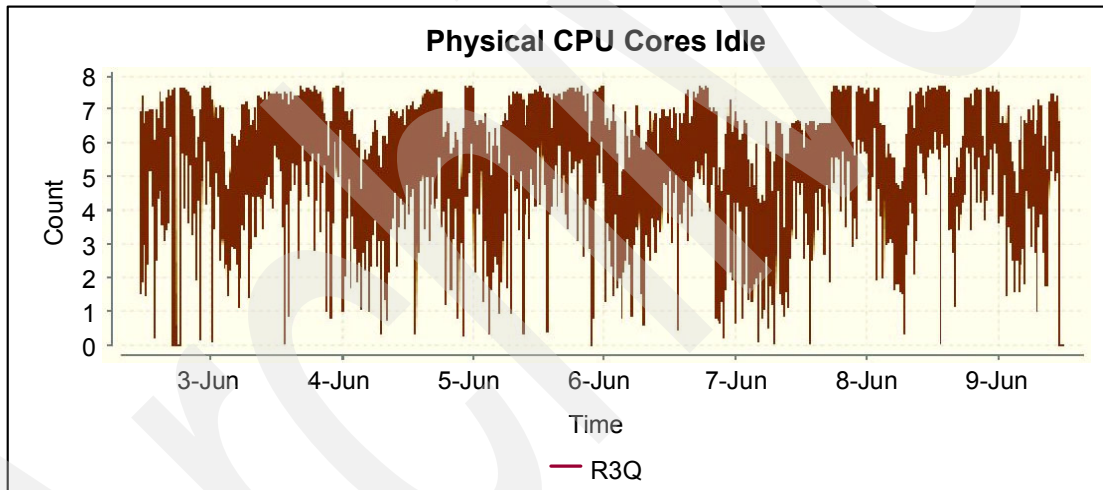
- ▶ Support for SAP's Central Monitoring System (CEN) in SAP ERP 6.0 and later or SAP NetWeaver environments (base component ECC 6.0 or later).
- ▶ Support for concurrent multi-SAP systems data collection (that is, the ability to collect multiple systems within one or many landscapes concurrently in an SAP NetWeaver environment).
- ▶ Capture and report distributed statistical records (DSR) enabling the monitoring and reporting of SAP Java stack components and statistics (for example, SAP Process Integration, SAP Enterprise Portal, SAP Application Server JAVA, and so on)
- ▶ Capture and report a customization measure of the SAP system (reported as a percentage of custom reports and transactions registered in the production SAP system catalog).
- ▶ Capture and report IBM virtualized system statistics as seen and reported by the SAP monitoring system.
- ▶ Enhanced and reorganized report format.
- ▶ Continued reporting of all previously published ABAP environment performance statistics.

Figure 9-39 on page 158 displays a sample page from an Insight Report.

## Host / LPAR Physical CPU Consumption – xs-bhm-srq-1 / xs-bhm-sr



Peak 1-hour average consumption of 1.2 CPU cores occurred on 06/08/2008 at 16:51.  
Peak 8-hour average consumption of 0.7 CPU cores occurred on 06/08/2008 at 19:03.



This chart plots physical CPU idle\* in one minute intervals over time.

Figure 9-39 Example page from an Insight Report

### 9.6.1 Installing and configuring IBM Insight for SAP

The IBM Insight for SAP utility program is packaged as an all-in-one Microsoft Windows executable. You can install it on any PC that has TCP/IP connectivity to the monitored production SAP system.

It does not need any additional third-party software on the collector PC or the monitored servers nor does it require any additional transports to be applied to the SAP system.

The Insight download packages and detailed information about the installation and configuration (including a must-read Readme file) are at:

<http://www.ibm.com/erp/sap/insight>



## 9.6.2 Data collection

After installed and configured with the details of the SAP system to be monitored, the Insight Collector is manually triggered to start and stop data collection, which is accomplished using RFC communication and is designed to make minimal impact on the monitored system. It is recommended that you collect at least one-to-three days of statistics during a peak period or a period of the month with a reasonably high usage to improve the quality and value of the report.

## 9.6.3 Report Generation Service

At the end of a data collection session, the Insight Collector can package the session data into a single compressed file that is ready to be emailed or FTP'ed to IBM. Analysis takes a minimum of three business days.

If you have questions about this utility, contact: <mailto:ibmerp@us.ibm.com>.

Archived



## Support statements by IBM and SAP

This chapter covers hardware and software requirements for the use of virtualization technologies in SAP landscapes on IBM PowerVM.

## 10.1 SAP general support statement

SAP has summarized the generic support for virtualization technologies in SAP Note 1492000. This is also a good starting point, because it contains references to several platform-specific SAP Notes.

## 10.2 AIX

SAP Note 1492000 contains a section that covers all AIX-specific SAP Notes related to virtualization.

## 10.3 IBM i

PowerVM virtualization is mostly transparent to the IBM i operating system, so you can refer to the SAP Product Availability Matrix for a list of supported SAP products and IBM i operating system releases. This is also valid for the database “IBM DB2 for i”, which is included in the operating system. There are a few cases where the SAP operating system monitor does not reflect certain virtualization technologies, including Active Memory Sharing (AMS) and Workload Capping. You may use these technologies, but you may need to get support from IBM rather than SAP in some circumstances of performance problems. In order to get support from IBM in those cases, it is highly recommended that you start an operating system performance data collection (see 9.2.2, “Performance data collection tools on IBM i” on page 111), and that you have the licensed program “IBM Performance tools for i” installed (57xx-PT1 - Manager Feature). For a support statement from SAP, see SAP Notes 1492000 and 1002461.

## 10.4 Linux

SAP supports the following IBM Linux distributions for use with SAP solutions:

- ▶ Novell SUSE Linux:
  - SUSE Linux Enterprise Server 11 (SLES 11)
  - SUSE Linux Enterprise Server 10 (SLES 10)
  - SUSE Linux Enterprise Server 9 (SLES 9)
- ▶ Red Hat:
  - Red Hat Enterprise Linux 6 (RHEL6)
  - Red Hat Enterprise Linux 5 (RHEL 5)
  - Red Hat Enterprise Linux 4 (RHEL 4)

Table 10-1 shows the support matrix of Linux distributions and processor generations.

Table 10-1 Support matrix

	POWER7	POWER6	POWER5
SLES11	✓	✓	
SLES10	✓ (SP3 and newer)	✓ (SP2 and newer)	✓ (SP2 and newer)
SLES9			✓

	POWER7	POWER6	POWER5
RHEL6	✓	✓	
RHEL5	✓ (QU5 and newer)	✓ (QU2 and newer)	✓
RHEL4		✓ (QU7 and newer)	✓

See also:

<http://publib.boulder.ibm.com/infocenter/lxinfo/v3r0m0/index.jsp?topic=/liaam/liaamdistros.htm> .

For more detailed information, see Chapter 2, “PowerVM virtualization technologies” on page 11.

To take advantage of POWER systems capabilities, such as Dynamic Logical Partitioning (DLPAR), you have to install additional software components. These are available and known as Diagnostic Tools for Linux Systems from the following website:

<http://www14.software.ibm.com/webapp/set2/sas/f/1opdiags/home.html>

Download the tools to the Linux system and install them according to the documentation on the website.

Archived

# Related publications

The publications that are 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 168. Some of the documents referenced here may be available in softcopy only.

- ▶ *IBM PowerVM Virtualization Managing and Monitoring*, SG24-7590
- ▶ *IBM PowerVM Virtualization Introduction and Configuration*, SG24-7940
- ▶ *Integrated Virtual Ethernet Adapter Technical Overview and Introduction*, REDP-4340
- ▶ *Advanced POWER Virtualization on IBM System p Virtual I/O Server Deployment Examples*, REDP-4224
- ▶ *Introduction to Workload Partition Management in IBM AIX Version 6.1*, SG24-7431
- ▶ *Deployment Guide Series: IBM Tivoli Monitoring V6.2*, SG24-7444
- ▶ *IBM PowerVM Live Partition Mobility*, SG24-7460
- ▶ *IBM PowerVM Virtualization Managing and Monitoring*, SG24-7590
- ▶ *Application and Program Performance Analysis Using PEX Statistics on IBM i5/OS*, SG24-7457
- ▶ *IBM iDoctor iSeries Job Watcher: Advanced Performance Tool*, SG24-6474

## Online resources

These websites are also relevant as further information sources:

- ▶ Workload Partitions (WPARs) and Shared Processor LPARs for SAP Environments  
<http://www.ibm.com/support/techdocs/atsmastr.nsf/WebIndex/WP101179>
- ▶ SAP's Adaptive Computing On IBM pSeries Implementation Guide  
<http://www.ibm.com/support/techdocs/atsmastr.nsf/WebIndex/WP100914>
- ▶ nmon for AIX & Linux Performance Monitoring  
<http://www.ibm.com/developerworks/wikis/display/WikiPtype/nmon>
- ▶ Micro-Partitioning statistics and documentation tool for IBM Power Systems & System i servers  
<http://www.ibm.com/developerworks/wikis/display/virtualization/LPAR2rrd+tool>
- ▶ Proposal to extend Ganglia with additional POWER5/6 metrics for AIX 5L and Linux  
<http://www.perz1.org/ganglia/>
- ▶ IBM Insight for SAP - Utility for anyone running SAP  
<http://www.ibm.com/support/techdocs/atsmastr.nsf/WebIndex/PRS381>

- ▶ Service and productivity tools - For Linux systems  
<http://www14.software.ibm.com/webapp/set2/sas/f/topdiags/home.html>
- ▶ AIX Wiki  
<http://www.ibm.com/developerworks/wikis/display/WikiPtype//AIX>
- ▶ Power Systems Virtualization Wiki  
<http://www.ibm.com/developerworks/wikis/display/virtualization/Home>
- ▶ Performance Capabilities Reference (SC41-0607)  
<http://www.ibm.com/systems/i/advantages/perfmgmt/resource.html>
- ▶ SAP Service Marketplace  
<http://service.sap.com>
- ▶ SAP Adaptive Computing  
<https://www.sdn.sap.com/irj/sdn/adaptive>
- ▶ SAP Help Portal  
<http://help.sap.com/>
- ▶ SAP Community Network  
<https://www.sdn.sap.com/irj/sdn/virtualization>
- ▶ SAP Notes are accessible, only with an S-user ID, at:  
<http://service.sap.com/support>
  - If you do not have such a user ID, right at the top of the Welcome window of the following website there is a link to request a new user ID if you are a customer or partner of SAP:  
<http://service.sap.com>
  - If you are an IBM employee, you can contact the ISICC Infoservice at the following website, and request an ID:  
<mailto:isicc@de.ibm.com>
- ▶ SAP Notes that will be of interest to you are:
  - 212876 The new archiving tool SAPCAR
  - 1002461 Support of IBM Dynamic LPAR and Micropartitioning
  - 1031096 Installing Package SAPHOSTAGENT
  - 1492000 General Support Statement for Virtual Environments
  - Live Partition Mobility:
    - 1102760 PowerVM Live Partition Mobility
  - WPAR:
    - 1087498 Support for AIX 6.1
    - 1105456 SAP Installations in AIX WPARs
    - 1131691 CPU utilization metrics of IBM System p
  - Memory Management:
    - 191801 AIX 64-bit with very large amount of Extended Memory
    - 445533 (for 4.6D and older) Lots of extended memory on AIX (64 bit)
    - 789477 (for 6.10 and later) Large extended memory on AIX (64 bit) as of Kernel 6.20
    - 856848 AIX Extended Memory Disclaiming
    - 1088458 AIX: Performance improvement for ES/SHM\_SEGS\_VERSION = 2



- Monitoring:
  - 994025 Virtualized OS environments in the operating system
  - 1019567 Corrections new operating system monitor
- IBM i
  - 45335 AS/400: Priorities of work processes
  - 49201 IBM i: Main storage settings for SAP
  - 808607 IBM i: Memory management in a system based on PASE
  - 863821 IBM i: Installing SAPCAR
  - 1023092 iSeries: Using Separate Memory Pools for Java Processes
  - 1123501: System i: Operating system user for SAP on System i
  - 1306787 IBM i: Standalone installation of CCMS agent
- ▶ You can request the following documents by contacting ISICC Infoservice at:
  - <mailto:isicc@de.ibm.com>
  - I<sup>2</sup>V - Intelligent IT Virtualization, Study by BearingPoint
  - I<sup>2</sup>O - Intelligent IT operations with IBM dynamic infrastructure solutions in SAP environments, Study by BearingPoint
  - A Banking IT Service Center runs SAP applications in a fully virtualized IBM PowerVM environment, SPC03286-DEEN-02
  - Automotive Supplier customer, technical reference story: Advantages of IBM POWER technology for mission-critical environments, GE12-2885-00
  - Global energy leader streamlines business applications and IT infrastructure with SAP and IBM, SPC03343-GBEN-00
  - Ciba technical customer reference story: Ciba demonstrates how to implement a complex, highly available SAP landscape without wasting processing capacity, using IBM virtualization and database technologies, GK12-4318-00
  - Stadtwerke Düsseldorf technical customer reference story: System landscape at Stadtwerke Düsseldorf on a virtualized IBM POWER environment, GM12-6851-00

## Documentation about Live Partition Mobility

- ▶ Live Partition Mobility in POWER6 Systems Hardware Information Center at:
  - <http://publib.boulder.ibm.com/infocenter/powersys/v3r1m5/index.jsp?topic=/iphc3/iphc3kickoff.htm>
- ▶ Live Partition Mobility in POWER7 Systems Hardware Information Center at:
  - <http://publib.boulder.ibm.com/infocenter/powersys/v3r1m5/index.jsp?topic=/p7hc3/iphc3kickoff.htm>
- ▶ Virtual fiber channel in POWER6 Systems Hardware Information Center at:
  - <http://publib.boulder.ibm.com/infocenter/powersys/v3r1m5/index.jsp?topic=/iphc3/iphc3kickoff.htm>

## White papers about Live Partition Mobility

- ▶ Live Migration of Power Partitions Running SAP Applications
  - <http://www.ibm.com/support/techdocs/atsmastr.nsf/WebIndex/WP101471>
- ▶ IBM Power Systems Live Partition Mobility (LPM) and Oracle DB Single Instance
  - <http://www.ibm.com/support/techdocs/atsmastr.nsf/WebIndex/WP101566>

- ▶ SAP & IBM Power Systems DLPAR and LPM Customer Demo  
<http://www.ibm.com/support/techdocs/atsmastr.nsf/WebIndex/WP101600>

## Other resources

- ▶ Setting Up Network To Improving Partition Mobility Performance  
<https://www.ibm.com/support/docview.wss?uid=isg3T7000117>
- ▶ DB2 and the Live Partition Mobility feature of PowerVM on IBM System p using storage area network (SAN) storage  
<http://www.ibm.com/developerworks/aix/library/au-db2andpower/index.html>
- ▶ JS22 Live Partition Mobility article on IBM developerWorks  
[http://www.ibm.com/developerworks/aix/library/au-js22lpm/index.html?S\\_TACT=105A GX20&S\\_CMP=EDU](http://www.ibm.com/developerworks/aix/library/au-js22lpm/index.html?S_TACT=105A GX20&S_CMP=EDU)

## How to get IBM Redbooks

You can search for, view, or download IBM Redbooks, Redpapers, Technotes, draft publications and Additional materials, as well as order hardcopy Redbooks, at this website:

[ibm.com/redbooks](http://ibm.com/redbooks)

## Help from IBM

IBM Support and downloads

[ibm.com/support](http://ibm.com/support)

IBM Global Services

[ibm.com/services](http://ibm.com/services)

# Index

## A

- activate and customize the Central Performance Histor  
146
- Activate and modify memory compression 38
- Active Memory Expansion 24, 27
- active Memory Expansion for SAP systems 84
- Active Memory Sharing 23
- Adaptive Computing 76
- Adaptive Computing compliance test 77
- Adaptive Computing concept
  - computing 76
  - control 77
  - network 77
  - storage 77
- Ad-hoc Performance Tools on IBM i 108
- AME 84
- application servers 1
- application WPARs 71
- assign implicit capping to LPARs 33

## B

- backup 54
- BASE 26

## C

- CCMS 127
- CCMS agent 144
- CCMS transaction OS07n 128
- CEC 29
- CEC view 104
- Central Performance History 146
- CFGPFRCOL 111
- characteristics of WPARs 70
- client LPAR 50
- client LPAR to the Virtual Network 52
- COLPRF 111
- Combination of workloads 28
- commands
  - CRTPFRSUM 111
  - DSPWLCGRP 84
  - license -accept 48
  - lparstat 100
  - lsmmap 50
  - mkdev 50
  - mktcpip 53
  - mpstat 99
  - ppc64\_cpu 41
  - QPRCMLTTSK 41
  - sar 98
  - smtctl 41
  - STRPFRCOL 111
  - STRPFRT 115
  - topas 103

- topasout 106
- topasrec 106
- viosbr 54
- vmo 40
- vmstat 80
- WRKSYSACT 110
- WRKSYSSTS 108
- compatibility modes 40
- compliance test 77
- concurrent load prioritization 8
- configure the RFC connection for the CCMS agent 144
- connecting a client LPAR to a virtual SCSI server adapter 50
- connecting a client LPAR to the Virtual Network 52
- consolidate 4
- consolidation 4
- consolidation effect 28
- CPH 147
- CPU time charged 86
- CPU time in SAP Applications 92
- CPU time measurement and SMT 87
- CPU utilization metrics 85
- create reporting on the Central Performance History 153
- create the report definitions 153
- creating
  - report definitions 153
- creating Virtual Ethernet Adapters 52
- creating virtual SCSI server adapters 48
- CRTPFRSUM 111
- customize the Central Performance History 146

## D

- Data collection 159
- dedicated LPARs 16
- dedicated to shared 29
- define how long and how granular data should be stored in the CPH 147
- defining the VIOS LPAR 48
- Demo of Live Partition Mobility with running SAP systems 68
- Downloading and installing Ganglia 125
- Dynamic LPAR 16
- dynamic LPAR operations 34
- dynamic work processes 78

## E

- Enable Pool Utilization Authority 38
- ENDPFRCOL 111
- Enhanced tools for root cause analysis 116
- Enterprise Edition 11
- ES/TABLE = SHM\_SEGS 78, 80
- evolution trends in the SAP landscape 1
- example at a virtualization customer 27
- examples of virtualization advantages 4

execute the report 155  
existing virtual adapters 50  
Express Edition 11  
extended memory tuning recommendations 78

## F

Flexible License Mechanism 67  
from a non-virtualized to a virtualized infrastructure 1  
Fundamentals of SMT 89

## G

Ganglia 121  
Ganglia and System p virtualization 124  
Ganglia metric terms 124  
Ganglia terms 122  
    cluster 122  
    grid 122  
    node 122

## H

Hardware Management Console 14  
high availability 47  
high availability for virtual networks 47  
high availability for vSCS 45  
how granular data should be stored in the CPH 147  
how long and how granular data should be stored in the CPH 147  
Hyper-Threading 89  
Hypervisor 13

## I

I<sup>2</sup>O - Intelligent IT operations 167  
I<sup>2</sup>V - Intelligent IT Virtualization 167  
IBM i 82, 167  
IBM i Disk Watcher 116–117  
IBM i Job Watcher 116–117  
IBM iDoctor for IBM i 117  
IBM Insight for SAP 157  
IBM PowerVM and SAP ACC Integration demo 68  
implementation example at a virtualization customer 27  
implicit capping to LPARs 33  
information about existing virtual adapters 50  
installation of SAP NetWeaver 7.0 in a Shared System WPAR 73  
Installing and configuring IBM Insight for SAP 158  
Installing in a non-VIOS LPAR 139  
Installing in a VIOS LPAR 142  
installing the VIOS 48  
instance profile changes 78  
Integrated Virtual Ethernet 23  
Integrated Virtualization Manager 14

## K

Keys 67

## L

landscape consolidation 4  
License 67  
Linux on Power 161  
    Red Hat Enterprise Linux 4 (RHEL 4) 162  
    Red Hat Enterprise Linux 5 (RHEL 5) 162  
    SUSE Linux Enterprise Server 10 (SLES 10) 162  
    SUSE Linux Enterprise Server 9 (SLES 9) 162  
Live Application Mobility 25, 74  
Live Partition Mobility 16, 63  
    documentation 167  
    whitepaper 167  
Live Partition Mobility Demo 68  
load prioritization 8  
load shifting 5  
LPAR 16  
LPAR2rrd 117  
lparstat 100  
LPM 63

## M

Main storage pools 82  
management of multiple Shared-Processor Pools 36  
management tasks 29  
memory compression 38  
Memory Management 166  
Memory Mapped to File 79  
memory tuning recommendations 78  
metric CPU time 91  
MGTCOL 111  
micropartition design option 32  
Micro-Partitioning and Shared Processor LPARs 17  
mktcpip command 53  
Monitoring 127, 167  
monitoring 95  
monitoring of virtualized SAP systems 127  
monitoring tools 117  
monitoring with OS tools 96  
motivation for server virtualization 4  
moving from dedicated to shared 29  
mpstat 99–100  
Multiple Shared-Processor Pools 18  
multiple Shared-Processor Pools 36

## N

N Port ID Virtualization 20  
nmon 106  
non-virtualized 1  
NPIV 20, 54

## O

Online resources 165  
OS07n 128

## P

parameter --smt 41  
Partition Suspend / Resume 19

- Performance 66
- Performance Data Collection Tools on IBM i 111
- Performance Explorer 116
- performance metrics that are to be stored in the CPH 151
- Performance Monitoring on IBM i 108
- PEX 116
- POWER6 to POWER7 Migrations 68
- PowerVM editions 13
- PowerVM virtualization technologies 11
- Prerequisites 65
- prerequisites for using PowerVM technologies on IBM POWER Linux 163
- processing chain 7
- Processor Utilization Resource Register 89
- PUA 38
- PURR 89
- PURR based metrics in POWER7 91

## Q

- QAPMLPARH 113
- QAPMSHRMP 113
- QMPGDATA 111
- QPFRRDATA 111

## R

- Red Hat Enterprise Linux 4 (RHEL 4) 162
- Red Hat Enterprise Linux 5 (RHEL 5) 162
- Red Hat Enterprise Linux 6 (RHEL6) 162
- Redbooks Web site 168
  - Contact us xiii
- report definitions 153
- Report Generation Service 159
- Resume 19
- RFC connection 144

## S

- sample installation SAP NetWeaver 7.0 in a Shared System WPAR 73
- SAN 54
- SAP Adaptive Computing 76
- SAP EarlyWatch 156
- SAP EarlyWatch and SAP GoingLive check services 156
- SAP general support statement 162
- SAP GoingLive 156
- SAP instance profile changes 78
- SAP landscape 1
- SAP License Keys 67
- SAP Monitoring Infrastructure 136
- SAP NetWeaver 7.0 in a Shared System WPAR 73
- SAP Notes 67, 166
- SAP Release upgrade and manual migration scenarios 73
- SAP support 71
- SAP technologies taking advantage of virtualization 75
- saposcol 131
- saposcol and CCMS agent 138
- sar 97
- SBSD 26

- schedule dynamic LPAR operations 34
- SCSI server adapters 48
- select the performance metrics that are to be stored in the CPH 151
- server virtualization 4
- setting up a VIOS partition 48
- Shared Dedicated Capacity 18
- Shared Ethernet Adapter 22
- Shared Processor LPARs 17
- Shared System WPARs without resource control in a dedicated LPAR 72
- Shared-Processor Pools 18, 36
- Side effects of PURR based metrics 94
- Simultaneous Multithreading 25, 40
- single core with SMT 89
- single core without SMT 87
- single statistic records 92
- SMT 40
- SMT on CPU utilization metrics 85
- SMT-2 40
- SMT-4 40
- SMT-4 mode 93
- smtctl 41
- Special considerations on IBM i 136
- ST03 93
- st06 127
- Standard Edition 11
- start of saposcol 144
- STRPFRT 115
- subsystems 26
- Support statements 161
- supported PowerVM technologies on IBM POWER Linux 163
- supported scenarios 72
- SUSE Linux Enterprise Server 10 (SLES 10) 162
- SUSE Linux Enterprise Server 11 (SLES 11) 162
- SUSE Linux Enterprise Server 9 (SLES 9) 162
- Suspend 19
- System i Navigator 114
- system WPARs 70
- Systems Director Management Console 15
- Systems Director VMControl 15

## T

- TCP/IP address for the VIOS 53
- Tivoli Monitoring 126
- tms\_utime 86
- topas 103
- topas -C 104
- trends in the SAP landscape 1
- tuning in AIX for SAP systems 80
- types of WPARs 70

## U

- unshared System WPARs without resource control in a dedicated LPAR 72
- Using single statistic records for application performance analysis 93
- using virtual SCS 45

## **V**

VCPU 29  
Versioned WPARs 71  
VIO Server 54  
VIO Server monitoring 54  
VIOS 20, 43  
VIOS backup 54  
VIOS LPAR 48  
VIOS partition 48  
Virtual Ethernet 21  
Virtual Ethernet Adapters 52  
Virtual I/O Server 19, 43  
Virtual I/O Server - basic usage types 44  
Virtual Memory tuning in AIX for SAP systems 80  
virtual network 46  
Virtual SCSI 20, 45  
Virtual SCSI Server Adapter 50  
virtual SCSI Server Adapters 48  
Virtual Tape 20  
virtualization advantages 4  
virtualization technologies 11  
virtualized infrastructure 1  
vmo 40  
vmstat 96  
vSCSI 45

## **W**

work process priorities 82  
Work process priorities and workload capping 83  
Work with System Activity 110  
workload capping 82  
Workload Partition Manager 25  
Workload Partitions 25  
workload partitions 69  
workload sizing 94  
WPAR 69, 166  
WPAR characteristics 70

## **Z**

zoning 54











# SAP Applications on IBM PowerVM



**Business  
responsiveness  
through granular  
flexibility of resources**

IBM invented the virtualization technology starting in the 1960s on the mainframe, and the functionalities evolved and were ported to other platforms and improved the reliability, availability, and serviceability (RAS) features. With virtualization, you achieve better asset utilization, reduced operating costs, and faster responsiveness to changing business demands.

**Transparent  
monitoring of the  
virtualized  
environment right  
within SAP CCMS**

Every technology vendor in the SAP ecosystem understands virtualization as slightly different capabilities on different levels (storage and server hardware, processor, memory, I/O resources or the application, and so on). It is important to understand exactly what functionality is offered and how it supports the client's business requirements.

**Leading  
enterprise-ready  
virtualization  
technology**

In this IBM Redbooks publication, we focus on server virtualization technologies in the IBM Power Systems hardware, AIX, IBM i and Linux space and what they mean specifically for SAP applications running on this platform.

SAP clients can leverage the technology that the IBM Power Systems platform offers. In this book, we describe the technologies and functions, what they mean, and how they apply to the SAP system landscape.

**INTERNATIONAL  
TECHNICAL  
SUPPORT  
ORGANIZATION**

**BUILDING TECHNICAL  
INFORMATION BASED ON  
PRACTICAL EXPERIENCE**

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

**For more information:  
[ibm.com/redbooks](http://ibm.com/redbooks)**

SG24-7564-01

ISBN 0738436046